

Why Harvard needs Summers

The head of Harvard University leaves much to be desired in terms of tact and demonstrable respect for those who disagree with him. But the university should stick with him, at least for the time being.

There are many reasons to argue that Larry Summers, the blunt-spoken head of Harvard University, should step down. But there are more compelling reasons for the president of one of the world's premier academic institutions to remain in place than for him to pack his bags.

Summers' immediate troubles flared up in January after he made controversial comments suggesting that differences in intrinsic ability might partly explain why so few women reach top-tier positions in science. Time will tell whether his comments wreak serious damage on the institution's reputation and its ability to recruit female students and staff. But the resulting uproar tapped a vein of deep and broad staff discontent about Summers' governing style — discontent on which this publication reported earlier this year (see *Nature* 433, 190–192; 2005). Anger peaked at two fiery staff meetings in February, one of which Summers described as “searing”. But it now seems that he will be able to ride out the current storm.

The economist and former US Treasury secretary has brought a corporate management style to Harvard's decentralized and scholarly environment. Aggressive, argumentative and domineering are just a few of the words used to describe him by foes and fans alike. The onus is now on Summers to learn some tact and to show greater respect for the diverse and gifted scholars who drive his institution. At a recent faculty meeting, he pledged to listen to staff more and to temper his words and actions.

Summers' past actions don't inspire much confidence in this regard. Within months of taking over at Harvard in 2001, he became embroiled in an unseemly public row with Cornel West, the institution's best-known African-American academic, who subsequently

departed for Princeton. And he provoked a backlash from academics who felt that he had failed to consult them sufficiently on ambitious plans for the university's expansion to a new campus in Allston, Boston.

Ultimately, however, Summers' presidency must be judged by his overall contribution to Harvard and by the academic advances it makes during his tenure. Many staff members, including scientists, believe that the positive contributions he has already made, and the changes he is planning, count for more than his controversial style or his public comments on women in science.

For one thing, Summers has become a champion of scientific research at Harvard. He wants the university to invest heavily in scientific facilities and to establish a better environment for interdisciplinary research. He wants greater collaboration and synergy between research groups and departments than some academics would naturally adopt. And he expects scientific institutions to form the core of the university's massive expansion into the Allston campus.

The success of these projects will not become clear for some time, but many of the researchers involved are brimming with enthusiasm. Even some of Summers' critics acknowledge that it is better to have a strong-minded president who makes bold decisions for the good of the university than a faint-hearted head lacking ambition. And should Summers resign in the current circumstances, it is highly unlikely that his successor would display the strong leadership skills that Harvard requires.

Summers, however, will need the active support of his staff if he is to drive positive change. Whether or not that support is forthcoming will become clear in the next few months. If it is, Harvard can prosper under Summers' leadership. ■

In pursuit of balance

Sunbelt states that have boomed economically should eventually earn a larger slice of the research pie.

In the past quarter-century, economic and political clout in the United States has shifted markedly towards the south and the mountain west. Yet research dollars remain heavily concentrated in other parts of the country (see page 10). This creates an imbalance that political leaders and university presidents in states such as Florida, Texas and Arizona are now working energetically to address.

When the existing structure of US research came into being after the Second World War, the lion's share of grant funding flowed to the most powerful universities of that era — almost all of them on the east coast, in the midwest and in California. Since then, the distribution of economic activity and population in the country has shifted. But the most powerful research departments are still concentrated in the same places.

To a large extent, that is as it should be. The greatest strength of the US system is its meritocracy, with grants distributed on the basis of a robust peer-review system. Another is its diversity: there are at least a dozen agencies that support significant amounts of university research. And the system has a third strength, commonly misdiagnosed as weakness: the occasional willingness of Congress to fund

specific projects directly. Used selectively, this process corrects the natural tendency of research dollars always to accumulate in the same places. Over time, these strengths will fuel the growth of top-flight research in the sunbelt states. For governors such as Florida's Jeb Bush, such growth can't come quickly enough.

There are many ways to encourage it. The best place to start is perhaps by establishing a robust public university. Arizona has fared well in this regard. But the more fragmented systems in Florida and Texas have struggled to attain the status to which the third- and fourth-largest states in the union must aspire.

Then there's the ‘big bang’ approach, as exemplified by Governor Bush's plan to invest a cool half-billion dollars in a branch of Scripps Research Institute at Palm Beach. There is some scepticism about the likely benefits of this plan, with critics saying that too many regions are banking on biotechnology as an economic engine. But Scripps Florida is not without logic. The original Scripps in La Jolla, after all, laid the foundation of the University of California at San Diego — now a research powerhouse. It is surely right for Florida and other states to aim high, and to aggressively pursue a larger slice of the science pie. ■





Word search

Physicists in a quandary over missing magazines
p4



Body of evidence

Hobbit remains prove a thorny issue for fossil hunters
p5



Welfare state

Japan debates rules to restrict animal use in the lab
p6



You've got male

Radio collars see sons rise for water voles
p8

NIH workers see red over revised rules for conflicts of interest

Meredith Wadman, Washington

The US National Institutes of Health (NIH) is facing a revolt by employees over its tightened rules on conflicts of interest.

The restrictions, which dramatically alter a policy set up in 1995, were announced on 1 February (see *Nature* 433, 557; 2005). They are the result of intense pressure on the NIH from an angry Congress, after a series of embarrassing disclosures revealed that a number of senior agency scientists had been making hundreds of thousands of dollars as consultants for commercial firms.

But the move has angered NIH employees, who say that the rules go too far. They point to talented young researchers who will be discouraged from coming to the agency because their inventions have been taken up by biotechnology companies (see 'Caught between a rock and a hard place', below). And they fear for long-standing employees whose life savings could be hit hard by regulations on owning stocks.

"Many of the rules are frightening," says Abner Notkins, chief of experimental medicine at the National Institute of Dental and Craniofacial Research. "They've gone to a damaging extreme."

"The vast majority of NIH employees have done nothing wrong. We are all being punished for the transgressions of a few," adds Elaine Jaffe, chief of haematopathology at the National Cancer Institute.



Scene of conflict: revamped ethical guidelines at the NIH are being criticized as too restrictive.

Under the rules, all forms of paid and unpaid consulting for biomedical companies are banned, and there are numerous restrictions on teaching and serving on company boards. Senior scientists are not allowed to accept academic prizes worth more than \$200. The rules also require some 6,000 senior NIH employees to sell any stock in biomedical companies owned by themselves, their spouses or their children by this July. The NIH's other 11,500 employees are

each limited to \$15,000 of stocks in any given biomedical firm.

Protesters say that the rules will hurt the agency's recruiting and retention of scientists, and force significant financial losses on many employees, especially those who are past the age of easily finding another job. Notkins points out that people who bought shares in Pfizer a year ago at \$37, for example, would be obliged to sell them at the current price of \$27.

Raynard Kington, NIH deputy director,

Caught between a rock and a hard place

When Elaine Jaffe read the new ethics rules for NIH employees announced last month, she immediately thought of the promising young physician she was seeking to employ in her lab at the National Cancer Institute in Bethesda, Maryland.

The physician was so bright that he had already mastered the first two years of coursework by the time he entered medical school. An engineer by training, he used his spare time at medical school to invent a machine that automatically embeds, processes and sections pathology specimens. He licensed it to a small biotechnology company for whom he worked

part-time while doing his medical residency. A patent is pending.

In the meantime, he applied for a two-year stint as a clinical fellow in Jaffe's haematopathology lab. He was enthusiastic about the chance to broaden his training, Jaffe says. Then the new NIH ethics regulations were announced.

They bar both temporary and permanent NIH employees from consulting for companies, whether or not they are paid for their services. Coming to the NIH would mean that the researcher couldn't be involved in steering the development of his invention. He is now reconsidering Jaffe's offer of a position.

Jaffe says that she understands why. "He won't be able to continue to work with the company to see the fruition of his effort. If they called him and they had a question about something and wanted him to take time off and come and work on something for a couple of days, he wouldn't be allowed to do it. It would stifle him creatively."

And yet, she says, "he wouldn't be working in my lab on anything related to his invention".

Jaffe says that she agrees with consulting restrictions on senior NIH scientists. But the ban for temporary employees is, she thinks, too much.

Meredith Wadman

counters that the rules are tough but necessary. "The preponderance of the evidence suggested that our ethics oversight system didn't work," he says. "Our number one priority was to ensure the public's trust in the integrity of the science of this agency."

He points out that, like it or not, NIH scientists can have an impact on financial markets—as they did in December, when Pfizer shares fell sharply after the National Cancer Institute halted a clinical trial amid safety concerns about the company's painkiller Celebrex.

Kington and other NIH officials have been probing the cases of some 100 NIH scientists who congressional investigators found were not complying with the old ethics rules. It has emerged that at least half of these people did not violate the rules, but were mistakenly identified because, for example, they had the same name as another researcher who had consulted for a drug firm.

"Even if 80% of them are cleared," says Kington, "having 20% on that list who may have violated the rules says something about the system."

But hundreds of NIH employees say that the agency's response is too extreme. In an online vote last month, 700 of them elected an executive committee for the Assembly of Scientists, a dormant group of intramural scientists that has reconstituted itself to try to soften the new rules.

The revival was spearheaded by Ezekiel Emanuel, chairman of the NIH Department of Clinical Bioethics, and quickly led to a two-hour meeting late last week between the assembly's executive committee and officials including agency director Elias Zerhouni and Kington. The assembly is also consulting the American Civil Liberties Union to see whether the new rules violate the privacy and freedom-of-speech rights of NIH employees.

In the meantime, the reality of the changes is becoming apparent on the NIH campus. Last week, officials at the National Cancer Institute circulated a call for nominations for the \$50,000 Paul Marks Prize for Cancer Research, offered by the Memorial Sloan-Kettering Cancer Center in New York. "Federal employees ... could accept the honor and the plaque, but not the monetary prize," the e-mail noted.

Kington said last week that the NIH is committed to assessing the impact of the new rules on recruitment and retention, and making modifications if "an appropriate evidence base" shows that they are necessary. He asked NIH employees to "provide their input as clearly as possible". Comments from the public are being accepted until 3 April at ethics@hhs.gov.

Opinion

Is There Really a Cowboy Culture of Arrogance at Los Alamos?

Brad Lee Holian

The director of Los Alamos National Laboratory (LANL), G. Peter Nanos, shut down all classified operations at the lab in mid-July, following a security incident in which two items of classified removable electronic media were reported to be missing after an inventory. Then, two days after a student intern apparently suffered an eye injury from a laser in a chemistry laboratory, the director shut down all operations at the lab, even though preliminary investigations had only begun. In an all-hands meeting, he told the LANL staff, "There's a belief amongst some very powerful people in Congress that academic culture and running a high-security national laboratory are totally incompatible, and scientists can't be trusted." He pointed

security, and compliance—often going so far as to suggest that the lab is a den of cowboy scientists, saboteurs, spies, and thieves—it is appropriate to consult the available statistics for the laboratory in these areas. One can then judge the behavior of workers at LANL compared to large industries and other nuclear weapons labs in the US Department of Energy complex, such as Lawrence Livermore National Laboratory and Sandia National Laboratories.

The accompanying figure shows timelines of recordable safety incidents (injury and illness) over the past decade, normalized to the equivalent of 100 person-years worked, for DOE lab (LANL, LLNL, SNL, ORNL, PNL, and YAL).

Why then has LANL been such a target of congressional scrutiny and ire? One answer might be that since the bombing of Hiroshima and Nagasaki by nuclear weapons built at Los Alamos, the lab has been the poster child of ban-the-bomb activism from the political Left, and from the lab's inception there have been calls to close it or at least terminate its nuclear weapons activity. From the political Right come the angry questions: "Why is there such a thing as 'academic freedom' at all in a laboratory? Why are any federal

Paper chase: copies of a contentious article were never delivered to Los Alamos staff.

Physicists miss out on critical points as magazines vanish

Geoff Brumfiel, Washington

Some sensitive material has once again disappeared from Los Alamos National Laboratory in New Mexico—and the corridors are abuzz with theories about what happened to it.

Gone missing this time are a couple of hundred copies of the December 2004 issue of *Physics Today*, which contained an article critical of Peter Nanos, the laboratory's director.

After many physicists at the laboratory reported that they had not received the issue, conspiracy theories began to circulate about the fate of the lost magazines.

Laboratory officials categorically deny that they are trying to keep the article from the staff. "The notion that there was some sort of an effort to keep *Physics Today* out of the hands of subscribers is ludicrous at best," says Jim Fallin, the laboratory's chief spokesman.

But according to an e-mail survey published in the March issue of the magazine, more than half of the laboratory's 414 subscribers say that they never received their copies of the December issue. By comparison, less than 3% say they didn't get the February 2005 issue.

The mystery of the missing magazines is just the latest reported disappearance at the laboratory: in July of last year, for example, two hard drives containing classified data were reported missing (see *Nature* 430, 387; 2004). The disks, together with the injury of a summer intern, led Nanos to shut down parts of Los Alamos for nearly six months and to accuse lab scientists of a "cowboy culture" of disregard for safety and security

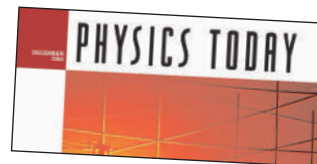
rules at the facility. Many researchers were livid about the long shutdown, especially after a government investigation concluded that the missing disks never actually existed (see *Nature* 433, 447; 2005).

The *Physics Today* article was an opinion piece by Brad Lee Holian, a theoretical physicist at the laboratory. It attacked Nanos's position by citing statistics that showed Los Alamos's safety record to be comparable to that of other national

laboratories (*Physics Today* 57(12), 60–61; 2004). So when the issue in which it was printed failed to arrive, physicists at the lab began to speculate as to the cause of the disappearance.

"The whole thing is a mystery," says Holian. He adds that he, for one, doesn't buy the conspiracy theories: "I'm still in the camp that somebody goofed up somehow." But, he points out, the fact that rumours continue to circulate on e-mail and the popular 'LANL: The Real Story' blog shows the level of tension between staff and administrators. "After the shutdown, people don't have a lot of confidence in laboratory management anymore," he says.

Fallin doubts that Los Alamos's staff or its management are to blame for the vanishing magazines. "Mailroom employees pride themselves in their handling of *Physics Today*," he notes sternly. A thorough search of the lab's mail facilities has yet to turn up the missing issues, and he says a more likely explanation may be that *Physics Today*'s mailing labels contained errors. The laboratory has contacted the postmaster-general in Albuquerque to help investigate.





Small skull, big discovery: *Homo floresiensis* (left) represents a new branch of human evolution.

Fossil finders in tug of war over analysis of hobbit bones

Rex Dalton, Jakarta

The prized bones of a miniature hominin have finally been returned to the scientists who discovered them, after months of dispute with a competing scientist who had taken them away.

The move is being seen as a victory by the discovery team. But some samples have yet to be given back. And a quarrel over whether the find really represents a new species continues to haunt the researchers.

"It is a complete circus," says Peter Brown, an Australian palaeoanthropologist who co-led the Indonesian–Australian team that reported the discovery last autumn^{1,2}.

The skeletal remains are those of *Homo floresiensis*, a metre-tall hominin species nicknamed 'hobbit', whose discovery revealed a new branch of the human evolutionary tree. The bones, found on the island of Flores, Indonesia, were dated to the unexpectedly recent time of just 18,000 years ago.

Within a month of the publication, one of Indonesia's top anthropologists, Teuku Jacob of Gajah Mada University, had taken the bones to his lab for analysis. These included the skeletal remains of eight individuals, some of which have yet to be described officially. Jacob was given access to the fossils by his friend and co-leader of the discovery team, archaeologist Radien Soejono of the Indonesian Centre for Archaeology in Jakarta.

The rest of the discovery team, also led by archaeologist Michael Morwood of the University of New England in Armidale, Australia, was furious that Jacob had removed the remains.

Tensions built as Jacob began saying publicly that *H. floresiensis* was not a new human species. He contends that the bones are from

Homo sapiens pygmies. The one dwarfed skull could be explained by a congenital defect, Jacob says. Many other leading palaeoanthropologists, including Tim White of the University of California, Berkeley, and Chris Stringer of the Natural History Museum in London, disagree with Jacob's interpretation.

The dispute worsened when Jacob gave two other researchers access to the bones for a week in February. Morwood and Brown call the examination of samples about which they have not yet published "unethical". But the researchers in question — Alan Thorne, a semi-retired anthropologist from the Australian National University in Canberra, and anatomist Maciej Henneberg of the University of Adelaide — say they only looked briefly at these specimens and deny any improper conduct.

Jacob promised to return the bones in both January and February, by deadlines agreed with the Indonesian Centre. But he twice failed to do so, saying that he had not finished with the remains. On 23 February, the bones were at last returned to the centre, where they are being held under lock and key.

But some samples remain elsewhere. Pieces of rib bone given out by Jacob for genetic analysis are still at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and at another lab in Jakarta. If successful, analysis of DNA should help to pin down the evolution of the species.

The discovery team is demanding that this material also be returned immediately. In the meantime, the researchers are in Indonesia looking for more bones to verify and expand their theories. ■

1. Brown, P. et al. *Nature* **431**, 1055–1061 (2004).

2. Morwood, M. J. et al. *Nature* **431**, 1087–1091 (2004).

France lays plans for premier cancer centre in Toulouse

Alison Abbott, Toulouse

Europe's largest centre for cancer research is to be built on the site of a major chemical explosion in France.

More than 30 people died and thousands were injured on 21 September 2001 in a blast at the Toulouse chemical company AZF. The explosion was one of Europe's worst industrial accidents.

The cancer centre will form the heart of one of France's seven new 'Cancéropôles' — regional networks of cancer research and care.

The 220-hectare site will house many cancer-research groups and institutes that are currently scattered throughout the city, as well as some hospital departments. These will benefit from shared facilities, including sophisticated genomics and microscopy equipment. An institute for advanced techniques in life sciences, employing physicists, chemists and computer scientists, will also be built.

The entire project is being directed by Georges Delsol, a cancer researcher at Purpan University Hospital, Toulouse. "The explosion was not only a terrible human tragedy, but it also threatened the long-term economic well-being of the city," says Delsol. "The Cancéropôle will bring new jobs as well as being very good for science."

Planned public investment in plant and equipment at the site will total about €600 million (US\$800 million). Two French pharmaceutical companies — Sanofi-Aventis and Pierre Fabre — will also move to the site, and say that they plan to build labs to develop new drugs, partly on the basis of the cancer research that will also take place there.

But first the land must be cleaned up. The explosion occurred in a store containing hundreds of tonnes of chemicals for fertilizer production. Total, AZF's parent company, is organizing and financing the clean-up of the pollutants — predominantly ammonia, ammonium nitrates and chlorine. It is selling the land to the Cancéropôle for the symbolic price of 'one franc'.

The ministry of defence, which still owns adjacent land used as a military dump, is cleaning that up and will also sell it cheaply to the Cancéropôle.

Building work will start in earnest next year, and the facility should be opened in 2008, when it will employ up to 2,500 people. ■

Japanese call for more bite in animal rules

David Cyranoski, Tokyo

Animal-welfare activists are locking horns with researchers in Japan over how strictly the use of animals in the country's labs should be controlled.

Friction has built up because the Japanese parliament is preparing an updated version of the 1973 animal-welfare law, which is expected to pass by June (see *Nature* **430**, 714; 2004). Animal-welfare campaigners want to introduce legally binding restrictions on experiments, but researchers advocate voluntary guidelines instead.

Mounting public concern has put pressure on parliament to strengthen legislation. The current system is based on "guesswork" about what is happening in laboratories, Seichi Kaneda, a member of the main opposition party, the Democratic Party of Japan, told a hearing in Tokyo on 24 February.

At the moment, general guidelines are enforced by the environment ministry on the basis of the 1973 law. But ethical questions relating to specific experiments, such as whether animals need to be used in the way proposed by the researchers, are dealt with by the ministries funding the research.

Researchers acknowledge that the system could be improved. A representative of the Science Council of Japan (SCJ), a coalition of scientific societies, told the hearing that Japan has a reputation "as an outlaw country without rules". The representative — a neuroscientist who asked not to be identified for fear of harassment by animal-welfare groups — said that Japan should implement



Guarded optimism: campaigners hope parliament will take action.

voluntary guidelines that would cover industrial as well as academic labs.

The guidelines proposed by the SCJ would require institutions to set up committees to monitor animal experiments. They would also call for greater use of third-party accreditation of labs by bodies such as the US-based Association for Assessment and

Accreditation of Laboratory Animal Care International.

The SCJ says that if these guidelines were in place, there would be no need for legally binding regulation of animal experiments, which it claims might be implemented arbitrarily by local government officials.

But animal-rights groups, such as All Life in a Viable Environment (ALIVE), say that self-regulation by researchers is not enough. A survey carried out by ALIVE last September showed that most university medical departments don't even keep track of the number of animals used. "They just want to avoid any kind of accountability," says ALIVE director Fusako Nogami.

The revised law should at least contain a registration system that requires experimental facilities or researchers to inform a government authority that an experiment will be carried out, says Nogami.

Most other countries have much stricter regulations. Britain requires labs to be licensed if they do animal experiments. The United States combines legal measures (including unannounced visits that can result

in fines) with voluntary guidelines.

Japan's Democratic party hopes that negotiations with members of the ruling coalition will produce a more restrictive draft of the law. Momentum for reform is apparently strong, although the civil service and the powerful Liberal Democratic party seem to be less enthusiastic about such restrictions. ■

Agency to bring fast-breeder reactor out of mothballs

Ichiko Fuyuno, Tokyo

Japan is taking steps to restart its prototype fast-breeder nuclear reactor, which has been sitting dormant in the northern state of Fukui since an accident in 1995.

In February, the Fukui government approved plans to restart the Monju reactor, effectively giving the go-ahead for a government plan to spend ¥15 billion (US\$140 million) over the next few years to get the prototype running again.

Critics of the move point out that fast-breeder technology has been abandoned in Britain, the United States and Germany.

"It was a mistake to build the Monju reactor. It is too expensive and time-consuming," says Hitoshi Yoshioka, a specialist in the history of science and science policy at Kyushu University

and a member of Japan's Atomic Energy Commission.

However, India and China are forging ahead with fast-breeders, which use mixed oxides of plutonium and uranium as fuel and produce more fissile material than they consume. Japan's determination to press on with the project suggests that in Asian nations with limited sources of energy the technology still has a future.

"Japan needs to look for new energy resources," says Hiroshi Nunota, an official at the nuclear fuel cycle section of the education ministry, which is funding the project.

The government of Fukui had been reluctant to allow the reactor to restart following public concern about an accident in 1995, when liquid sodium coolant leaked

from the reactor's secondary cooling system.

Critics of the project claim that the state only reversed its decision after the central government agreed to provide economic packages and to link Fukui to Tokyo by bullet train.

Japan's Nuclear Cycle Development Institute has been working on the project since 1968 and its cost to date is estimated at ¥800 billion. A final obstacle to its resuming operation — a 2003 court ruling that nullified its construction permit — is likely to be overturned during the next couple of months, say energy analysts.

The institute still needs to win approval from the local government and residents before restarting the operation. But most analysts think that this approval will also be forthcoming. ■



Write stuff: is work on *Escherichia coli* being ignored in the scramble to fund bioterror research?

Protest letter accuses health agency of biodefence bias

Erika Check, Washington

Hundreds of US biologists have signed a letter protesting at what they see as the excessive use of bacteriology funds for the study of bioterror threats.

The letter, which reflects growing unease among researchers, was due to be delivered this week to managers at the National Institutes of Health (NIH), US lawmakers and the leaders of seven scientific societies.

"The diversion of research funds from projects of high public-health importance to projects of high biodefence relevance represents a misdirection of NIH priorities and a crisis for NIH-supported microbiological research," the letter states.

Its 750 signers include two Nobel laureates and seven past presidents of the American Society for Microbiology. The protest was organized by molecular biologist Richard Ebright of Rutgers University in Piscataway, New Jersey. In the letter, Ebright writes that the National Institute of Allergy and Infectious Diseases (NIAID) awarded 15 times more biodefence grants between the beginning of 2001 and the end of 2004 than it awarded during the previous four-year period.

Meanwhile, Ebright says, the agency cut grants to study non-biodefence models, such as *Escherichia coli*, by 41% and grants to study non-biodefence microbes that cause disease by 27%. He argues that this shift is preventing important advances in science and public health, and actually increases the risk of a bioterrorism incident.

"Bioweapons agents cause, on average, zero deaths per year in the United States, in contrast to a broad range of non-prioritized microbial pathogens that cause tens or hundreds of thousands of deaths per year," Ebright says.

Not just funders but investigators are shifting their focus to biodefence-related microbes, it seems. "We have become unbalanced," says Martha Howe of the University of Tennessee in Memphis, a past president of the American Society for Microbiology. Researchers are just not making the basic-science applications, she explains.

But Anthony Fauci, director of the NIAID, says he disagrees with the premise of the protest. "Although I have a great deal of respect for the people who signed that letter, if they understood all the issues and numbers involved I don't think they would be as concerned," he says.

Fauci cites the NIAID's own data, which show that the agency funded about the same amount of research in basic bacteriology in 2000 as in 2004. Over that period, Fauci says, the number of grants increased slightly, from 131 to 137, and never dropped below 120, while the funding dropped very slightly, from \$40,741,867 to \$40,502,815, after hitting a low of \$34,168,719 in 2003. Fauci says his data are more appropriate than Ebright's, because they track all awards made across the institute.

"Ebright's saying that biodefence is taking away from non-biodefence infectious disease and microbiology," says Fauci, "and the facts clearly show that is not the case." He adds that grants awarded outside biodefence have dropped across the entire NIH owing to tight fiscal restraints in recent years. Just last week, Fauci said that the institute might also have to cut AIDS research.

But for Ebright, that is simply proof that biodefence spending is damaging other research. "The main constraint that is placing pressure on all other components is the biodefence budget," he says. ■

Pasteur researchers win fight to stay in city centre

Declan Butler, Paris

Scientists at the Pasteur Institute in Paris have won a long running battle over plans to move labs to a commercial zone on the outskirts of the city. In a report presented to the management on 24 February, external arbitrators said that the move was "not necessary".

The mediators were John Skehel, a virologist and director of Britain's National Institute for Medical Research in London, and John Wills, the UK institute's administrator. They said that the Pasteur's plans to renovate its labs in central Paris could be accomplished in phases, without moving staff elsewhere.

Philippe Kourilsky, the director-general of the Pasteur Institute, had planned to move hundreds of scientists to a new site at Fresnes, southeast of Paris (see *Nature* 432, 788; 2004).

Staff protested that the site was in an undesirable area with poor public transportation. A petition signed by more than half the staff says that they "understand neither the necessity, nor the rationale" of the move.

The report highlights staff concerns that there should be "minimum effect on the progress of interactive research programmes during refurbishment". Additional space could be created, it adds, by temporarily moving BioTop, Pasteur's biotechnology 'incubator unit', to one of Paris's many science parks.

"The mediators' conclusions are a complete disavowal of management," says Agnès Labigne, head of Pasteur's Pathogenesis of Mucosal Bacteria unit.

Stewart Cole, senior vice-president for scientific affairs, says management will take the report's advice. "Skehel and Wills did a serious job; they came here six times, and had complete freedom to go anywhere, and to speak to anyone," he says. "Their conclusions are clear."

Although scientists seem likely to stay on campus during the refurbishment, a broader question remains regarding the need to expand to a second or new campus — to build high-throughput biology platforms, for example. Cole says that this will be given "more thought".

The mediators' report emphasized that any expansion should be firmly based on a long-term scientific strategy worked out in full consultation with the institute's researchers. "What we need now is reconciliation," says Cole. "The report provides us with the framework to do this." ■

Pressure group sues food agency for not adding salt

Washington A health-advocacy group is suing the US Food and Drug Administration (FDA) for failing to classify dietary salt as a food additive.

The Washington-based Center for Science in the Public Interest says there is scientific evidence that salt raises blood pressure and prematurely kills 150,000 Americans a year. But the FDA classifies dietary salt as “generally recognized as safe”, giving the agency no authority to limit how much salt foods can contain. The advocacy group says it hopes the lawsuit will force the FDA to change salt’s status.

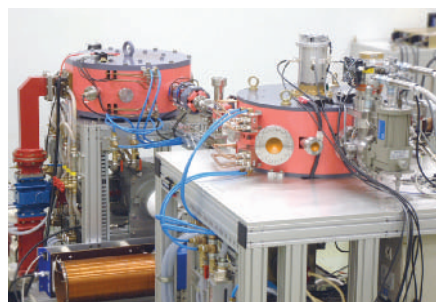
Americans consume about 4 grams of salt a day — roughly twice the amount recommended in the US government’s recent dietary advice (see *Nature* 433, 794–799; 2005). Overindulgence occurs in other countries too. The UK government is currently waging a campaign to cut Britain’s salt intake by one-third.

Future looks bright for table-top synchrotron

Tokyo A Japanese company has developed a synchrotron light source small enough to fit in a laboratory — although the handy technology comes at a hefty price.

Synchrotrons produce X-rays that can be used to probe the structure of materials. Their unwieldy size means that scientists must travel to large facilities and wait their turn to use the rays for their experiments.

Researchers have come up with theories on how to shrink the device (see *Nature* 428, 789; 2004), and now a working miniature is available for sale — the MIRRORCLE-6X, manufactured by the Photon Production Laboratory of Shiga, Japan. With a storage-ring diameter of 60 centimetres, the whole machine easily fits in a lab, and generates X-rays of up to a few mega-electronvolts using a novel electron-injection technology. Large facilities, whose synchrotrons are tens to hundreds of metres across, typically



Downsized: the MIRRORCLE-6X weighs in at under half a tonne and easily fits in the lab.

Sex changes seen in radio-collared voles

London The use of radio-transmitter collars to track and monitor animals has dramatically skewed the sex ratio of a population of endangered water voles, British ecologists have found. They fear the technology may be further endangering this population, and could have similar effects on other studied animals.

The scientists, led by wildlife researcher Tom Moorhouse of the University of Oxford, UK, began fitting a Norfolk population of *Arvicola terrestris* (pictured) with radio collars three years ago to study their migration and mating behaviour. Over the course of the project, they observed a 48% decline in the expected number of female offspring from tagged animals (T. P. Moorhouse and D. W. Macdonald, *J. Appl. Ecol.* 42, 91–98; 2005). They think that stress may be responsible for the shift: voles raise more males in hard times, as they are more likely to survive.



Radio collars were thought to cause some stress, but few scientists believed that they would influence behaviour or reproduction. Moorhouse says he will stop using the collars after further investigating the effect.

produce rays of about the same energy.

At roughly US\$2.5 million a piece, the MIRRORCLE-6X is unlikely to find its way into most laboratories any time soon. But Hironari Yamada, who helped to develop the machine at the Ritsumeikan University in Shiga, says that private businesses — from semiconductor manufacturers to pharmaceutical producers — are lining up to buy the device.

Indian institute gets cash to aim for the top

New Delhi The Indian government has granted Rs1 billion (US\$23 million) to the Bangalore-based Indian Institute of Science to help it to develop into a “world class university”.

The surprise announcement was made by the finance minister Palaniappan Chidambaram while presenting India’s annual budget on 28 February. He said the institute is the first of many that will receive huge funding increases in coming years, as the government works to boost India’s international competitiveness.

“We were not expecting this generosity,” the institute’s director Goverdhan Mehta told *Nature*. “More than the money, what makes us really happy is the recognition that we deserve this.”

Enzyme washing powder cleans up rogue prions

London A new way of decontaminating medical equipment might reduce the risk of prions being transmitted to patients during surgery.

Steel surgical instruments are usually sterilized by washing and heating. But studies have indicated that this does not remove prions — abnormal proteins thought

to cause variant Creutzfeldt–Jakob disease, the human version of mad cow disease. The only method proven to rid instruments of prions is bathing them in corrosive chemicals. In the face of a tiny theoretical risk of prion transmission, the UK government in 2001 allowed surgeries to use tools that have been simply washed and heated.

Researchers at the UK Medical Research Council Prion Unit in London now say that using a biological washing powder, with added enzymes called proteases, does the trick (G. S. Jackson *et al.*, *J. Gen. Virol.* 86, 869–878; 2005). The team hopes that a fine-tuned version of the process will be available for use in surgeries by the end of the year.

Launch success lifts Japan’s space hopes

Tokyo Japan’s space agency is celebrating the successful launch and deployment of a weather satellite last Saturday that has put it back in the Asian space race.

The launch success comes as a relief after a string of recent failures, including the loss of an Earth-observing satellite blasted by solar flares, two spy satellites that had to be blown up after a botched launch, and a mission to Mars that ran out of fuel. “We could not afford another failure,” says science minister Naruaki Nakayama. The space agency says that improved reliability and reduced costs should help them compete with China for future commercial launches.

Japan’s rocket programme is unusual in that it is not an offshoot of a military missile programme. But observers note that it is increasingly being seen to have potential military applications. The country is now moving to work with the United States on a missile defence programme.

The shape of things to come: Florida hopes that the planned development of Scripps Florida at Palm Beach will provide a huge boost to the state's economy.

Upstart states

The United States has a settled arrangement for distributing its research budget around the country, and the same states have dominated it for decades. But, as Emma Marris discovers in Florida, the have-nots have had enough.

Seen from the air, Florida's Palm Beach looks like paradise. There's the aquamarine sea, the sandy line of beach studded with palms, the pastel cul-de-sacs with their deep-blue squares. It is only as you descend that you realize that the squares aren't swimming pools, but plastic tarpaulins — makeshift repairs after last summer's hurricane season. Welcome to the future home of Scripps Florida.

Scripps is the state's heavyweight contender in a battle to win promotion to the first division of science. Jeb Bush, the president's brother and governor of Florida, has helped to arrange \$310 million in state funding to attract the Scripps Institute in La Jolla, California, to the state. Palm Beach County has promised a further \$200 million, including a 777-hectare former orange grove as a site.

Will the arrival of Scripps succeed in establishing a premier biomedical research centre and nurturing a string of spin-off biotechnology companies in this playground for the rich, tanned and famous? It's not only a question for Floridians. Other states that fare badly in the carve-up of

federal research funds (see map, opposite) will be watching Florida's progress with interest. One way or another, they all want the answer to the same question: is it too late for them to get into the science game?

Prestigious start

The basic geography of research funding in the United States dates from the system's establishment after the Second World War. Scientists, credited with the development of the atomic bomb, found their prestige at an all-time high, and the agencies set up by federal government to support their work had no hesitation in awarding the lion's share of the money to places where the scientists wanted to live and work. Mostly that meant Ivy League institutions on the east coast, the strong public universities in the midwest, and Stanford and the University of California system in the west.

The amount of funds available has since expanded to some \$23 billion, which the federal government will this year spend on basic and applied research in universities. Hundreds of less-celebrated institutions have

been seeking a share of the spoils, as have state governments such as Florida's. A successful research university is increasingly seen as a route to economic development. States that have traditionally received little research funding hope that building such institutes and encouraging spin-off companies will create high-paying jobs and attract new companies.

"When a university gets money, the effect is not confined just to that university," says Daniel Greenberg, a Washington-based journalist and author who has been writing about research policy for decades. "When you have a university with a thriving chemistry department you might get a paint manufacturer who says: 'This is a good place to locate a plant, because we can get consultants down the road very easily'."

Most federal science funds are distributed by government agencies on a competitive, peer-reviewed basis. Despite occasional carping that the peer-review system acts as an 'old-boys' network', most analysts consider it to be in pretty good shape. In general, good science is done by the best scientists, who are

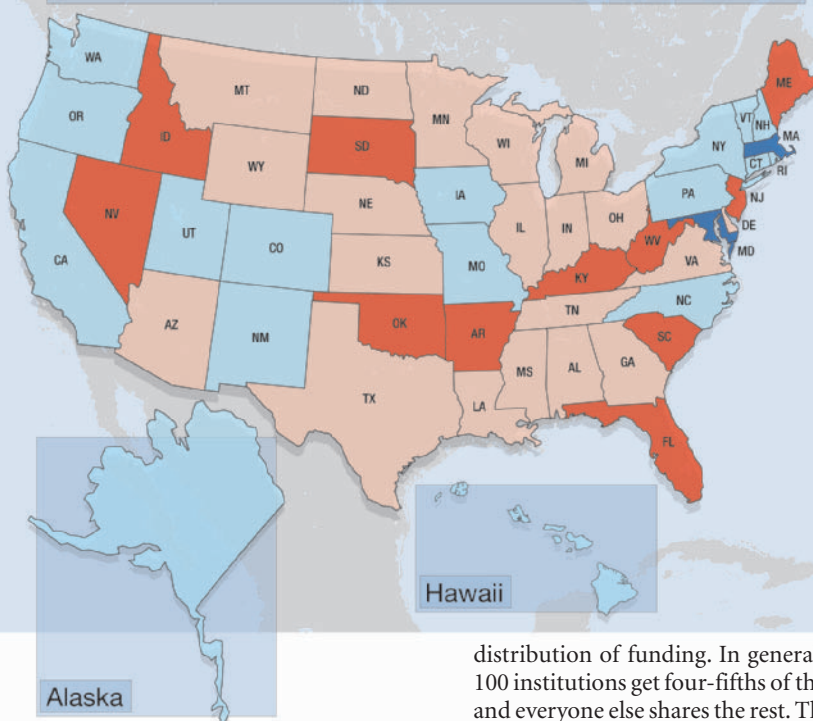
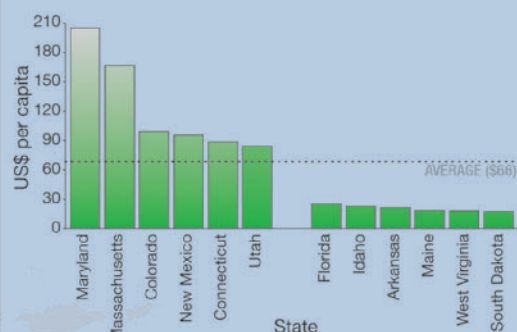
Federal spending on academic research and development

In 2000 (the year of the most recent census), the US federal government spent an average of \$66 per head of population on academic research and development. But this figure varied widely from state to state: in Maryland the spend was \$206 per person, in South Dakota, it was just \$18.

SPENDING PER CAPITA, 2000



WINNERS & LOSERS



attracted to the best institutions in search of the best research environment. It's a fair system — but a self-perpetuating one. States such as Texas, Florida and Arizona, which have enjoyed massive growth in their populations and economies during recent years, are still struggling to make their mark in science.

The National Institutes of Health (NIH) has come to dominate the research scene since its budget doubled to \$27 billion in the five years to 2003, and it now accounts for two-thirds of funding for academic research and development. But the agency's rise in funding has now flattened out. And some observers say that the boom, which benefited every academic research centre in the nation, has fostered delusions of grandeur in the 'have-not' states.

Big ideas

Joe Cortright, an analyst at Impresa, an economics consultancy based in Portland, Oregon, says that a study he did in 2002 for the Washington-based Brookings Institution showed a clear trend. Almost everywhere had twice as much biomedical funding in 2001 as in 1995, he says.

"What the governor hears is 'Our state doubled its funding.' This has led a lot of people to believe, naively, that their community is a big player," Cortright says.

Yet the boom hasn't radically altered the

distribution of funding. In general, the top 100 institutions get four-fifths of the money, and everyone else shares the rest. There is no sign that this will change as funding growth slows. Yet from coast to coast, Cortright notes, relatively obscure universities continue to spin off research parks left, right and centre, and local politicians are flush with money for scientific initiatives.

Florida is the fourth most populous US state, and its economy and population are among the fastest-growing in the nation. But in 2000 — the year of the most recent census — it ranked 44th out of 50 in a league table of the number of research dollars that academic institutions attracted per head of population. The state's best-known research facility is probably NASA's Kennedy Space Center, home of the space shuttle and assembly point for the International Space Station. The University of Florida advertises its science programmes as "more than just Gatorade", referring to the electrolyte-laced sports drink — the Gainesville university's most famous invention that has generated \$80 million in licensing income for the university.

The Scripps initiative is Jeb Bush's attempt to transform the state's scientific reputation in one fell swoop. He hopes that science can become a pillar of the state's economy, providing better-paid jobs than the current big employers: tourism, military bases and citrus farming.

Scripps would be the largest of a number

of science projects, from a small, existing medical-device industry to a 'high-technology corridor' stretching across the centre of the state from the Gulf of Mexico to the Atlantic and anchored by the University of Central Florida in Orlando and the University of South Florida in Tampa.

Attractive prospect

The Scripps project isn't the first attempt to kick-start high technology in the 'sunshine state'. In 1990, Florida unexpectedly beat Massachusetts to host most of a national magnet laboratory, a facility supported by the National Science Foundation. The National High Magnetic Field Laboratory (NHMFL) is now well-established, with one site tucked behind Florida State University at Tallahassee, another at the University of Florida in Gainesville, and a third at Los Alamos National Laboratory in New Mexico.

The NHMFL is a world leader in very powerful magnets. The Tallahassee site houses some of the world's largest resistive and hybrid magnets, which look like out-sized water heaters and use about 10% of Tallahassee's entire power supply. The lab is used by some 1,000 researchers each year in various disciplines to investigate molecules and materials.

Greg Boebinger, who joined the NHMFL as director last year, was as surprised as anyone when Florida won the laboratory from its previous site at the Massachusetts

Institute of Technology. "I was among the people who said: 'What is this?'" he admits, laughing. "One of the main reasons Florida got it was state support." The state put up about \$80 million to get the contract.

Boebinger is a great believer in the economic value of basic research. The magnet lab is like most scientific investments, he says: it repays into the community "an order of magnitude" more than it cost the state to build it. Federal funds for the NHMFL flow on into the local economy. An economic assessment conducted for the state last year by Florida State University's Center for Economic Forecasting and Analysis reported that Florida had got back three-and-a-half times its initial investment in the facility.

The NHMFL has also been mentioned by Scripps officials as a possible collaborator. Pat Griffin, head of drug discovery for Scripps Florida, visited the facility in January to discuss this idea. Florida officials are betting that Scripps can replicate the magnet lab's success on a far grander scale.

Sunny outlook?

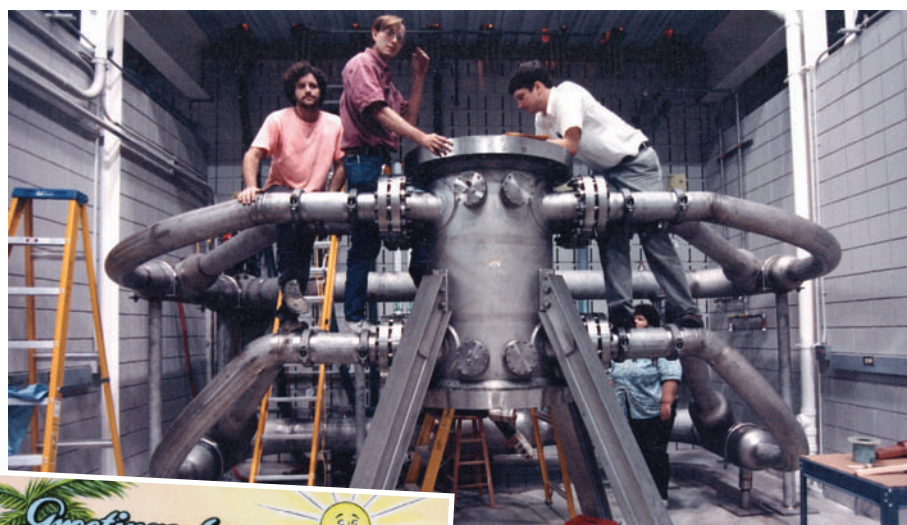
Last September, at a meeting of the economic development organization Enterprise Florida, several speakers argued that Scripps could act as the centrepiece of a coordinated plan for the development of the life sciences in Florida. Presentation after presentation forecast a rosy scientific future for the state. An analyst from Ernst & Young assured the audience, for example, that biotechnology in the United States was "clearly on the road to profitability for the first time"; and predicted it would be reach that goal in 2008.

But some speakers were more cautious. Kenneth Kirby, now president of TransDermal Technologies, said he had encountered difficulties in starting up his drug-delivery company in Lake Park, Florida.

He identified a funding gap for start-ups in the state, saying that venture capital there is relatively under-developed. Another speaker hit a nerve by joking that Florida universities, conditioned by years of competitive football, can't seem to collaborate very well.

Indeed, several competing universities would like to be champions of Florida's scientific future, and their rivalry could hold the state back, suggests Irwin Feller, an economist and science-policy specialist at the American Association for the Advancement of Science in Washington. Florida, he says, "fits the profile of a state where politics is in the way". Feller adds that local politicians tend to expect an unrealistically quick return on their investment. "All their interest is economic."

Sena Black, a vice-president of Enterprise Florida, notes that science-based companies started by researchers from outside the state can depart as soon as they become large enough to hire an experienced chief execu-



tive. "To get them to stay in Florida, they have to be Florida-bred, but Floridians are not being educated in science," she complains, adding that school science and mathematics education is weak. But she remains optimistic. "We have these pockets of science. It's more than meets the eye."

Digging deep

Now the state seeks a larger pocket of excellence. In October 2003, Palm Beach County spent \$60 million buying a large family farm for the Scripps site. What was once wetlands, and then rows of citrus trees, is now a muddy field surrounded by slash pines and palmettos. But lawsuits from environmentalists have put the choice of the site into some doubt, and Scripps may now consider at least one other site near Palm Beach for the complex.

Cortright isn't convinced by the ambitious Scripps Florida project. "I am extremely sceptical that it will produce any kind of biotech industry there. Just because they do research doesn't mean that companies will open up. The money they are spending on Scripps moves it from way, way, way below to just about where everyone else is," he says.

Griffin, who moved to Palm Beach from New Jersey to work at Scripps, is cautious too. "It's definitely not a guaranteed success," he says, "and I think the next few years will determine whether Scripps can do what it wants." In the meantime, he is enjoying living where he used to go on holiday.

Greg Schuckman, director of government

Ray of hope: Florida has already carved out a scientific niche by hosting the National High Magnetic Field Laboratory (above).

relations at the University of Central Florida, thinks that Florida should be able to exploit its reputation as a land of sun, beaches and easy living. He hopes that scientists can be lured from "the tundra of the midwest, the cost of living in California and the winters of the northeast".

Yet attempting to boost a regional economy with science is a risky move, perhaps riskier than some realize. Many state initiatives, including Florida's, are focusing on biotechnology. In Cortright's judgement, the game in this sector is over and the winners are already in: San Francisco, Boston and San Diego, plus pockets around the NIH in Maryland and at Research Triangle Park in North Carolina. "Anyone pursuing it now is throwing their money away," he thinks.

Feller is less dismissive. "If a region can pull together the faculty and provide the resources, the opportunities are there," he says. Larger states that are trying to punch their weight in science will only do so if local institutional rivalries within the states are kept at bay, he suggests. But history shows that it is possible for states with little scientific activity to work their way up into the system over decades. "The system isn't rigid," he says. "It's just very stable."

Greenberg generally agrees. He also points out that projects funded by Congress without peer review — known as 'earmarks' — can give a leg-up to institutions in the have-not states. These are generally absent from the National Science Foundation and the NIH, he says, but even these agencies have programmes aimed specifically at states that attract little peer-reviewed funding.

Still, any movement tends to be glacial in pace. "The country and its scientific enterprise are both very mature now," says Greenberg. "These are not the gold-rush days anymore."

Emma Marris is a reporter for *Nature* in Washington DC.

Reality check

They were highly prized artefacts with inscriptions that dated back to biblical times. The only problem was they were fake. Haim Watzman unearths the authentication work that has rocked Israel's archaeology community.

At the end of December, five alleged members of an antiquities forgery ring were indicted in Jerusalem's district court. At the same time, the Israel Museum in Jerusalem removed from view one of its most prized pieces: an ivory pomegranate bearing an inscription hinting that it had been used by priests in Solomon's Temple, the holiest site of the Israelite nation in the biblical period. The five men stand accused of forging this and other inscriptions.

Behind these events lies a debate between the museum and scientists over how and when archaeological objects should be authenticated. Should a rigorous, scientific, peer-review process have kicked into action over the pomegranate decades ago, when the piece was bought by the museum? Or were the expert opinions solicited by the museum at the time sufficient to justify its purchase — at a cost of US\$550,000?

Yuval Goren, chairman of Tel Aviv University's department of archaeology and ancient Near Eastern cultures, and head of the lab that recently investigated the pomegranate, bridges both sides of the debate at the heart of the problem. Is chemical and microscopic analysis the ultimate arbiter of authenticity, or can this be overruled by the work of historians and experts in ancient script? As an archaeologist with



Solomon's Temple is thought to have been sited where the Dome of the Rock now is in Jerusalem.

training in microscopy, Goren speaks the languages of both the natural and social sciences. At 48 years old, intense and good-looking, Goren could easily be the model for the hero of an archaeological detective series.

Goren and his team were first brought in to look at the pomegranate last year. It was of paramount importance to historians at the time, as it was thought to be a rare piece of solid evidence for the existence of Solomon's Temple. According to the Bible, this temple was built in the tenth century BC at the command of God ('Yahweh' in ancient Hebrew), on a hilltop that is now Islam's third holiest site, holding the Dome of the Rock and the Al-Aqsa mosque. The site's sanctity means that it can't be excavated, and the resulting dearth of direct evidence has led some historians, as well as adversaries of the modern state of Israel, to argue that Solomon's Temple never existed and that the Jewish state thus has no legitimate claim to the sacred hill.

The pomegranate's inscription, obliterated in part by a hole in the object, has been reconstructed to read: "Belonging to the Temple of Yahweh, holy to the priests," in ancient Hebrew. The cream-coloured relic has a hole in the bottom, and the archaeologists who examined it when it first came to light

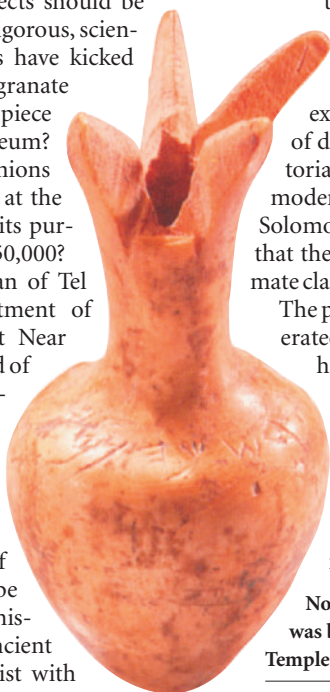
speculated that it was used as the top of a sceptre. Pomegranates, being one of the seven species with which God blessed the Holy Land, according to the Bible, are a common motif in Israelite artefacts.

When the Israel Museum bought the piece from an undisclosed owner in the 1980s, curators asked Nahman Avigad, a senior archaeologist at the Hebrew University of Jerusalem, to authenticate the object. Avigad and his colleagues examined the piece with a microscope and declared it to be genuine¹. But by 2003, serious doubts were raised about this conclusion when two other highly publicized objects with biblical associations threw up some surprises.

Genuine article?

The first of these was a black sandstone tablet, engraved with an inscription in ancient Hebrew remarkably similar to two biblical passages describing temple repairs carried out by King Jehoash of Judah. Three geologists at the Geological Survey of Israel examined this 'Jehoash tablet' in 2002 and determined, through chemical analyses and carbon dating, that the stone and its patina — the coating that often develops on ancient stone, metal and ceramics through interaction with air or soil — indicated that the inscription was more than 2,000 years old. They ruled that it was probably a remnant of Solomon's Temple².

But when Edward Greenstein, a professor of biblical studies at Tel Aviv University, saw a picture of the tablet in his morning newspaper early in 2003, he quickly came to the conclusion that it was a forgery, and not a very good one. He and other palaeographers



Not what it seems: this pomegranate was believed to come from Solomon's Temple, but it is now thought to be fake.

— experts in ancient scripts — said that some of the word uses were modern rather than ancient, and that in several cases the spelling wasn't right for the time period. "Any one of these howlers would demonstrate the spurious character of the inscription," wrote Frank Moore Cross, an emeritus professor of Hebrew at Harvard Divinity School in Cambridge, Massachusetts, in the *Israel Exploration Journal*³. If there was conflict between the natural-science results and the palaeographic results, Greenstein and Cross argued, the natural scientists were wrong.

The second object was an ossuary of the type used in Roman times to inter human bones, with an inscription reading "James son of Joseph brother of Jesus". In this case some palaeographers and archaeologists declared the item to be authentic whereas others claimed it was a forgery.

Out of character

More work on these objects was clearly needed. So in 2003, the Israel Antiquities Authority began an investigation. It established two committees, one of humanities experts and one (with Goren as a leading member) of natural scientists.

In the case of the Jehoash tablet, Goren issued a scathing critique of the geologists' methods and conclusions⁴. In the analysis he and his colleagues performed, they noted that the patina within the inscription's letters is quite different from that on the back of the tablet. The latter is packed with silicon and strongly attached to the rock surface, he noted, and so is undoubtedly authentic. But the patina in the inscription is a mixture containing chalk, iron-rich clay, charcoal and microscopic globules of gold — a composition that matches neither the tablet itself nor the rocks and soil of Jerusalem. Goren and his associates concluded that this patina was fake.

In the case of the James ossuary, Goren found that the patina in and around the inscription contains coccoliths — microfossils of tiny marine creatures. Such fossils could not have been transferred to the object from the chalk in the surrounding soil by natural processes, Goren says. Instead, he says, the forgers probably ground up chalk and mixed it into a fake patina.

Goren's collaborators also looked at oxygen isotope ratios in the patina, which provide information about the conditions, such as temperature and humidity, in which it formed. They found these ratios to be widely different in the letters themselves compared with elsewhere on the object. Goren and his two co-workers concluded that the patina on the inscription could not have been created in Jerusalem during the past 3,000 years: the inscription, he says, is a recent addition.

The only other possibility is that someone, for some reason, scraped off the real patina and replaced it. This provides an



Hollow words: careful examination has shown the inscriptions on the Jehoash tablet to be fake.

opening for those textual scholars who still maintain that the inscription is real.

In June 2003 the committees issued their reports: both objects were forgeries. Most of the humanities scholars agreed, although there are some who still reserve judgement.

Tracking the source

When suspicions about the objects first arose, police sought them out and found them in the possession of Oded Golan, a private antiquities dealer and one of the five indicted men, who continues to assert that the objects are genuine. He claimed that he was trying to sell the pieces on behalf of their owners, whom he did not identify. In 2003, the antiquities authority and the

police asked Goren and others to examine a number of other inscribed objects that had passed through Golan's hands. The Israel Museum set up its own task force, with Goren as a leading member, to examine the prize piece: the pomegranate.

Goren's lab is the opposite of high-tech, containing some microscopes and devices for measuring and cutting tiny portions of potentially valuable objects for study. In the case of the Jehoash tablet and the James ossuary, Goren scraped off tiny amounts of patina using a scalpel, a small chisel and a set of dental tools. He then used microscopic images to help determine the composition and crystalline structure of the object's minerals, revealing where the stone was probably



Fraud squad: Yuval Goren (below) led a team that concluded the James ossuary (above) featured a fake inscription.



quarried and the conditions under which the patina formed. If needs be, Goren's chemist colleagues can use instruments such as mass spectrometers to determine the exact elemental composition of the material.

Goren concluded that although the pomegranate does date to the bronze age — the period 3300–1200 BC and before Solomon's Temple is believed to have been built — its inscription is a modern addition. "The patina is no patina — it contains only silicon," Goren says. "Furthermore, whoever wrote the inscription was careful not to cross an ancient fracture in the object. That means the inscription postdates the fracture. It's hard to believe that someone would dedicate a

broken pomegranate to the temple," he says.

Michal Dayagi-Mendels, the Israel Museum's chief curator of archaeology, is not happy about the debunking of a prize object that cost the museum half-a-million dollars. But she says that this may be the sad but inevitable consequence of the advancing science of detecting frauds. "Tests become more and more sophisticated over the years," she says. "The kinds of tests that were done now could not have been done when we made the acquisition."

Letter by letter

But Goren disagrees. "The methods I use are classical methods that everyone uses. They're available and inexpensive," he asserts. True, he says, the use of the scanning electron microscope, a device that provides high-magnification, high-resolution images without damaging artefacts, is relatively new to archaeological investigations. But this just provides a new way of performing essentially the same analyses that have long been done on such objects. The same results could have been achieved in the 1980s with a conventional microscope, he says, just as his team did for the ossuary and the tablet.

Goren argues that forgers manage to dupe curators because museums are often so eager to gain a prize item that they take shortcuts through the scientific process. "It's not a problem of scientific method but of ethics," he says. As far as Goren is concerned, the work done on the pomegranate when it was purchased simply wasn't sufficient.

Dayagi-Mendels says that more than 90% of the archaeological artefacts owned by the Israel Museum come from documented excavations, which helps immensely in proving authenticity.

"The methods I use are classical methods that everyone uses. They're available and inexpensive."

— Yuval Goren

"In principle, anything that doesn't come out of an excavation undergoes a series of tests in our own restoration laboratories and, when called for, in external laboratories," she says. Such objects, as well as excavated objects about which questions arise, undergo further testing over the years, she adds. James Snyder, the museum's director and an art historian by profession, says the procedure followed for the pomegranate's purchase, which was made before his directorship, was perfectly appropriate.

Added value

It is not practical to run every possible test on every unprovenanced object, notes Paul Craddock, a materials scientist at the British Museum's department of conservation, documentation and science. Craddock, who is now writing a book on fakes and forgeries, emphasizes that the scientists can only seek to answer questions that curators pose for them, based on their expertise and professional intuition. He cites a case in which his museum was offered a bronze water vessel from the medieval period, which was unusually shaped as a unicorn. "We had to be told that the important bit was the horn. If we'd run our tests on a leg, the object would have been proved genuine," he says. But with the knowledge that it was the horn that made the object valuable and collectable, he ran an ultraviolet test over that. "It showed it was a separate piece that had been glued on," Craddock says.

With such guidance, it should be easier for people such as Goren to help sort out the real pieces from the fakes. But there will always be debates about authenticity. The pomegranate, for one, has now become symbolic of such disputes. Snyder and Dayagi-Mendels have decided to put it back on display at the Israel Museum to serve as a lesson to the public. "We plan to keep it on view and to use it to explain the process of authentication," Snyder says.

"What's interesting is that science is illuminating but it is never definitive," Snyder adds. "My guess is that in the months to come we will see questions raised about whether the latest conclusions about the pomegranate are definitive." With new chemical results battling against historical analyses, and in some cases an overwhelming desire to believe that an inscription is real, coming to a truly final conclusion may simply be impossible. ■

Haim Watzman is a freelance writer in Jerusalem.

1. Avigad, N. *Israel Mus. J.* 8, 7–16 (1989).
2. Ilani, S., Rosenfeld, A. & Dvorachek, M. *GSI Curr. Res.* 13, 109–116 (2002).
3. Cross, F. M. *Israel Exploration J.* 53, 119–123 (2003).
4. Goren, Y., Ayalon, A., Bar-Matthews, M. & Schilman, B. *Tel Aviv* 31, 7–14 (2004).

Final report of the Israel Antiquities Authority's expert committees

► www.bibleinterp.com/articles/Final_Reports.htm

How volunteering for an MRI scan changed my life

Discovering a serious problem not only causes shock but can have financial implications.

Sir—I read your news story “Brain-scan ethics come under spotlight” (*Nature* 433, 185; 2005) with great interest. As a neuroscientist, and being a bit of a ‘neuro-nerd’, I’ve always wanted to observe MRI scans of my own brain, so when the opportunity arose I jumped at the chance to help test a new MRI facility at my university.

As it turns out, I should have thought about the consequences of volunteering more thoroughly.

After the test scans, the manager of the facility informed me that something abnormal had been observed during the procedure. With great trepidation, I looked through the scans and, having taught neuroanatomy previously, I instantly recognised a tumour, roughly the size of a golf ball, in a rather sensitive location near the carotid artery to the left of my brainstem. This came as a huge surprise as I had never been diagnosed with any sort of neurological disorder.

Some would call this a fortunate

discovery, and I would normally agree with them. Clearly, knowing you have a brain tumour is better than not knowing, right? The manager of the MRI facility offered to refer me to a local neurosurgeon for further investigation. In a state of shock, I agreed without proper consideration. This decision, I later realized, would have unforeseen financial implications.

At the time, my wife and I were expecting our first child, and we were in the process of reviewing our insurance policies. We had decided to apply for additional insurance to support the family should one of us lose our university position through injury or disease. Just before we submitted these documents, along came this ‘diagnosis’.

The neurosurgeon told me that 5% of operations lead to potential complications after which, in order to save my life, they would have to induce a massive stroke of my entire left-brain. This could leave me in the horrible position of being unable to communicate with my wife, my newborn

child or my students. Clearly, this surgery could lead to my losing my job. What should I do about the insurance policy? Revise the application and report these ‘non-clinical’ scans? I decided to be honest (others would say naïve) and report the scans, which cost me the policy.

Now I sit in the uneasy position of facing surgery that could cost me and my family everything because I wanted to peep at my own brain. I understand that subject recruitment for research studies can be very difficult and every subject is precious. After my experience, however, I feel that informed consent should clearly include recognizing the possibility that something of medical significance could arise and that this could have an impact on future insurance eligibility.

Sadly, this is likely to further reduce subject participation in research critical to our understanding of the healthy and diseased brain.

Name and address withheld by request

Coping with unsuspected findings in volunteers

Sir—I read your News story “Brain-scan ethics come under spotlight” (*Nature* 433, 185; 2005) with a feeling of déjà vu.

Nearly 20 years ago, my colleague Alfredo Vazquez and I reported a group of ‘normal’ volunteers who were discovered to have serious abnormalities during the course of research.

Of the three apparently healthy young medical students concerned, one was found to have chronic persistent hepatitis, another had a tumour in the parietal lobe of the brain, and the third had positive HIV serology (M. Phillips and A. J. Vazquez *Control. Clin. Trials* 8, 338–342; 1987).

Everyone’s perceived roles suddenly changed, and the outcome was devastating. The subjects ceased to be healthy students with bright prospects. They were furious that the research had transformed them into patients under threat of death. The researchers ceased to be physicians in full control of the situation. They were transformed from investigators into counsellors, and that left them confused and defensive.

All the research studies had been approved by an ethical committee.

We suggested two ways to improve future research protocols: participants should be alerted in advance to the

possibility that the research in which they are participating might reveal a previously unsuspected illness, and investigators should be required to formulate a plan to cope with this contingency.

Maybe the time has come to act on this modest proposal?

Michael Phillips

Menssana Research, 1 Horizon Road, Suite 1415, Fort Lee, New Jersey 07024-6510, USA

Solid evidence for bubble fusion?

Sir—Your News story “Bubble-based fusion bursts onto the scene” (*Nature* 432, 940–941; 2004) states that bubble-fusion work “remains in limbo” after research conducted at Oak Ridge National Laboratory (ORNL) yielded inconclusive results in 2002.

However, a second paper was published by researchers at Purdue/ORNL in March last year and provided additional evidence for bubble fusion (R. P. Taleyarkhan *et al. Phys. Rev. E* 69, 036109; 2004). This paper underwent a second thorough review by a different group of ORNL scientists who supported its publication.

In my view, the 2004 paper provides evidence that shifts the question from “Can we drive fusion this way?” to “Can we produce net fusion energy this way?”

A great deal of work will have to be

performed before that more difficult question can be answered.

Ross Tessien

Impulse Devices, 13366 Grass Valley Avenue, Grass Valley, California 95945, USA

India must cooperate on tsunami warning system

Sir—Following the 26 December tsunami, international survey teams working in the affected areas (“On the trail of destruction” *Nature* 433, 350–353; 2005) have held educational seminars attended by government ministers, local professionals, emergency management and, on occasion, even students. The emphasis has been on explaining tsunami generation and impact, lessons learned from recent tsunamis, information on the operation of tsunami warning centres and preliminary findings.

Unlike similar meetings in Sri Lanka, the Maldives and Indonesia, the meeting hosted by the Indian National Academy of Sciences in Delhi on 21–22 January was more focused on presenting national capabilities in remote sensing, seismology and storm warnings. The meeting concluded with a list of action items. There was little discussion of arguably the most fundamental aspect of a warning system – the communication of the warning and actions resulting from this. This omission was supposedly justified by India’s experience with storm warnings.

Yet neither this storm-warning experience nor the existence of India's sophisticated seismic networks led to warnings being issued on 26 December, once the tsunami had struck the Andaman and Nicobar Islands. The tsunami did not hit the Indian mainland for another two hours. Allegedly, communications links had survived in Port Blair on the Andaman Islands, not to mention nearby airforce and navy bases that were affected. Some have argued that up to 40,000 people might have been saved if they had been warned. Further, India issued an incorrect warning a few days after the tsunami hit, triggering massive panic in India and Sri Lanka.

At the same meeting, India announced that it could develop new systems and models "based on end-to-end principles" in two years, using the best brains in India. For reference, the United States and Japan took more than 20 years to develop validated numerical models to predict tsunami evolution. And it took the US National Oceanic and Atmospheric Administration 30 years to fully develop its bottom-pressure recorders, which have been reliably detecting tsunamis for the past ten years.

India has an opportunity to establish a regional warning centre for the Indian ocean, thus ending its self-imposed isolation in sharing seismic data. It has the communications infrastructure and the scientific talent to serve its citizens and the international community. But the idea that India can do it alone is misguided.

Costas Synolakis

Department of Civil Engineering,
University of Southern California,
Los Angeles, California 90089-2531, USA

Ethics and ethnoflora

Sir — Pleased though we were to read the generous and thorough review of our book *Ethnoflora of the Soqatra Archipelago* in *Nature* ("Back to the roots" *Nature* **432**, 805–806; 2004), we feel that — given the politically sensitive environment in which we work — we need to address the comment that "many ethical issues ... are not generally considered in this book".

We were scrupulous, for example, about informing all those who contributed to our book about their intellectual property rights. This was in strict compliance with the Convention on Biological Diversity (and with our contract). All informants recorded on tape, in Soqotri, their agreement and their understanding of the purpose of the research.

Regarding the recognition of intellectual contribution, our book includes a list of some 140 Soqotran contributors. (Female informants could not be named individually for cultural reasons.) And

although *Ethnoflora* did not specifically mention any of the educational and capacity-building programmes in the archipelago in which we are involved, these activities are fully reported elsewhere.

Tony Miller*, **Miranda Morris†**

*Royal Botanic Garden, Inverleith Row,
Edinburgh EH3 5LR, UK

†School of History, University of St Andrews,
St Katherine's Lodge, The Scores,
St Andrews, Fife KY16 9AL, UK

Biologists do not pose a threat to deep-sea vents

Sir — Magnus Johnson suggests, in Correspondence ("Oceans need protection from scientists too" *Nature* **433**, 105; 2005), that "uncoordinated and unregulated" research is one of the greatest threats to hydrothermal vent habitats. We offer information to the contrary. Furthermore, we suggest that the vent-research community is unusually well-organized internationally to examine the effects of researcher activities and to implement a code of conduct.

As with most field studies, it is possible to cite examples of overexuberant sampling, especially in the years following the discovery of vents. But potential effects of sampling were recognized early (V. J. Tunnicliffe *Geophys. Res.* **95**, 12961–12966; 1990) and researchers at vents are proactive in developing mechanisms to reduce sampling effects.

Although it is true that the main effects on hydrothermal vents come from scientists because the only visitors at vents are scientists, today much more emphasis is placed on management and conservation to reduce the collection of organisms. Many known vents are no longer sampled and effort is concentrated at a few sites.

The current ethos of vent marine scientists is evident in the activities of the Biogeography of Chemosynthetic Ecosystems (ChEss) programme (www.soc.soton.ac.uk/chess) within the 'census of marine life' initiative. ChEss helped to convene a fact-finding workshop on hydrothermal ecosystems with the United Nations' International Seabed Authority (ISA) last September. The ISA is responsible for developing the legislation required to ensure and provide for responsible and sustainable activity throughout the world's deep-ocean environments. There is also a draft Code of Conduct pending approval by InterRidge (www.interridge.org), the office that coordinates international studies on mid-ocean ridges.

Canada's Endeavour Hot Vents Marine Protected Area, which Johnson highlights, was established with the strong support

of scientists. Examine the website that Johnson cites to see that there are 'zones' of activities — including 'No Sample' areas. Johnson's comment that a senior scientist advised him not to complain is a sad one. Any discipline needs to keep its ears open to possible abuse as well as ensuring responsible reporting of the facts.

We have worked as scientists on many aspects of deep-sea oceanography for nearly 30 years and share all concerns about damage to that environment. The lessons we have learned at hydrothermal vents are ones that we now apply at other chemosynthetically driven ecosystems, such as cold seeps and whale falls.

Paul Tyler, Christopher German,
Verena Tunnicliffe

Southampton Oceanography Centre, University of
Southampton, Southampton SO14 3ZH, UK

Signed on behalf of 18 international members of the ChEss
programme steering group

Making sure corrections don't vanish online

Sir — We have counted the numbers of errata, corrigenda, corrections and addenda published in all 2004 *Nature* print issues, across all sections of the journal. During the past year, *Nature* has reported flaws in 32 peer-reviewed research papers, of which 24 were corrigenda (author corrections) and 8 were errata (journal corrections); there were also 2 addenda. Although all these corrections were published last year, 14 of the erroneous papers were published before 2004. Within the other sections of the journal there were 14 errors reported.

Worryingly, in 14 out of 34 cases *Nature* failed to attach an amendment page with the online PDF of the original paper. *Nature* has also failed to provide an amendment notice with the abstract or HTML version of several research papers (7 times out of 34). Similar problems arise for corrections to non-research items such as News stories.

With the widespread practice of accessing, printing and circulating PDF files through the Internet, it seems advisable to take this matter seriously.

Eun-Hee Shim*, **Vishwas Parekh†**

*Department of Biochemistry,

†Department of Hematology-Oncology,
St Jude Children's Research Hospital,
Memphis, Tennessee 38105, USA

Corrections published in *Nature* should be linked online to the article being corrected, both in the text of the correction and in the HTML of the original article. The articles identified by Shim and Parekh have now been linked in this way. Editor, *Nature*



Will we be ready for the next one? Storm-surge waves, such as this one from Hurricane Eloise, which hit Florida in 1975, can be as deadly as tsunamis.

Watching over the world's oceans

A quick technological fix is not the best response to the December tsunami.

Keith Alverson

In the mid-nineteenth century, the HMS Beagle docked in Concepción, Chile, giving Charles Darwin the opportunity to see and describe the immediate aftermath of a tidal wave. His eyewitness account in the classic *Voyage of the Beagle* could easily be read as a report from Sri Lanka after the tsunami of 26 December 2004. The timeless nature of the devastation stands in stark contrast to the enormous progress that has occurred since then in relevant areas of science, technology and intergovernmental cooperation — progress that should have made a difference. Plate tectonics, accurate seafloor mapping, powerful computer calculations for wave propagation, real-time wireless global communications networks and operational 24-hour government warning systems are all new since Darwin's time. It seems they made no difference. With hindsight, they could have, and should have.

The December tsunami was a natural catastrophe, but much of the death and destruction that followed was a result of the collective failure of human institutions. Not surprisingly, hindsight has informed the global response. In addition to the outpouring of aid, there is interest from nations wishing to build an operational tsunami warning system in the Indian Ocean as soon as possible. Although laudable, this goal is far too narrow. Why? Despite local tsunamis being a frequent occurrence in the Indian basin, we have no idea when or where to expect the next large regional tsunami. It could be centuries away. A rapidly developed, single-basin, single-purpose tsunami warning system that

goes unused for many years is likely to be falling apart by the time it is called to use.

This is not a wholly pessimistic view — we have been here before. Following two major tsunamis in the Pacific in the early 1960s, the Intergovernmental Oceanographic Commission of UNESCO (IOC) and its member states set up a warning system for that ocean. By 2004, the funding for the upkeep of that system was a trickle, and three of its six seafloor pressure sensors were out of commission. There has long been talk of expanding and upgrading the Pacific warning system, which lacks regional tsunami warning centres in many vulnerable areas — southeast Asia, the southwest Pacific, and Central and South America. Unfortunately, once the initial system was in place, the resources required to maintain it properly — let alone expand or improve it — were extremely difficult to find. Building a single-use warning system for the Pacific basin alone in response to the events of the early 1960s was arguably not the best thing to do. It would be a mistake for the international scientific community to suggest another quick technological fix for the Indian ocean, where tsunamis are even less frequent.

A multihazard approach

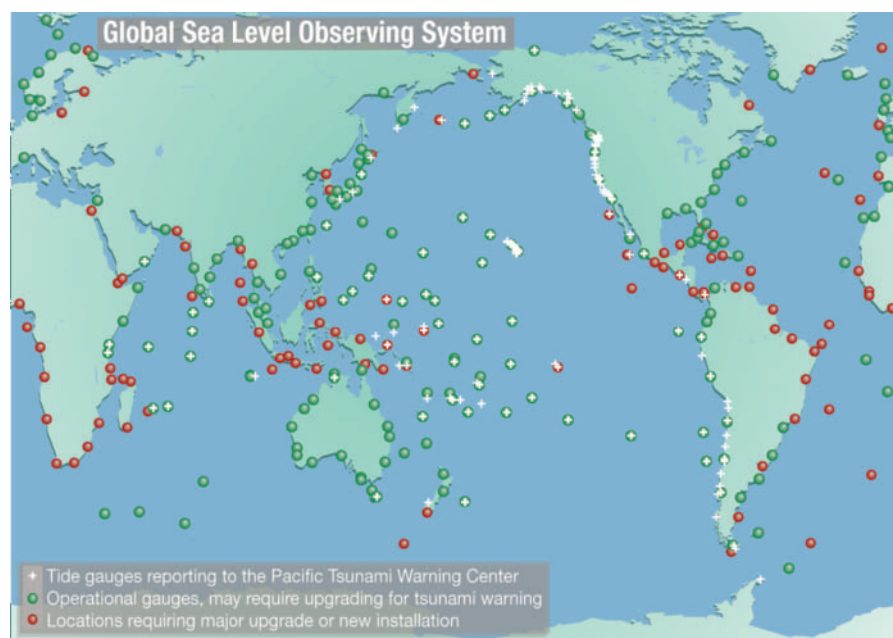
A more sensible idea is to develop a global tsunami warning system that is fully integrated with an operational ocean-observing system — one that is regularly used for other related hazards, such as storm surges. Storm surges associated with tropical cyclones can hit coastal areas well ahead of the landfall of the actual storm; they travel

with nearly the same rapidity as tsunamis, but occur much more frequently. And for unprepared or unwarned populations, they can be equally deadly. For example, in 1970 (and again in 1991) six- to seven-metre-high storm surges striking Bangladesh resulted in around half a million deaths. At present, there is no regional system for predicting storm surges, although there are a few national efforts. But tide gauges provide vital information for the high-resolution models used in storm-surge prediction — and these are the same data needed for tsunami warnings.

Although the scope of a tsunami warning system should be global, one of the most important components of any future network will be the national warning centres. Japan, Chile, New Zealand, Australia, French Polynesia, United States and the Russian Federation already run operational tsunami warning centres 24 hours a day, seven days a week. The track record of these centres is substantial, but it is time to improve the scope of their activities by working to build an operational, global ocean-disaster warning, preparedness and mitigation system.

In addition to detecting multiple hazards — from storm surges to cyclones — the best way to ensure that a tsunami warning system remains fully operational for decades to come is to embed it in broader efforts to observe the ocean. Data used for tsunami warnings are of potential interest to an enormous array of users and stakeholders. It is these other users who will ensure the system is maintained over the long term.

For example, changes in observed sea



level occur across many timescales, from seconds and minutes (wind waves, tsunamis), hours to days (tides, storm surges), and years (seasonal cycles, El Niño), through to long-term changes associated with climate change and the movement of land masses. Ocean circulation and long-term sea-level trends are monitored by the global array of tide gauges maintained by the Global Sea Level Observing System (GLOSS; see above) a component of the Global Ocean Observing System (GOOS). These are both run by the IOC, which aims to build a network of roughly 300 sea-level stations around the world (100 more than there are now), as well as several higher density regional networks.

Although some GLOSS stations already glean and process data in real time for the Pacific Tsunami Warning System, they operate mainly to serve the research community in a delayed mode. Upgrading the GLOSS network to real-time data delivery would contribute to a global tsunami warning system, and at the same time vastly increase its usefulness for other purposes.

For example, real-time sea-level data could contribute to ocean models serving a wide spectrum of users — including captains of large tankers who need predictions for efficient route planning. In such contexts, these data are of substantial economic interest. They can aid ship piloting in harbours, the management of sluices and barrages, tidal predictions and computations for coastal engineering design and insurance purposes.

The way forward

There are three substantial hurdles that need to be overcome to achieve this vision. The first challenge will be to develop an operational 'real time, all the time' capability for the ocean observing system. Those components of GOOS most relevant to marine hazards,

such as sea surface temperature, and sea-level and seafloor pressure, need to be made available in real time. This is not just a technical requirement, but also a difficult political issue. For example, some countries purposely limit the release of public data to monthly mean sea-level values, years after the fact, whereas their high-frequency data (1–2 minute averages) are kept private for reasons ranging from cost to national security. In addition, national centres running operationally 24 hours a day, seven days a week, are essential to a hazard warning system. With the exception of a few countries, oceanography does not have the required institutional support at the national level to enable such operations, and creative solutions will be required.

The second challenge will be to bring together the different scientific communities, such as seismologists involved in tsunami warnings, meteorologists involved in storm-surge warnings, and oceanographers involved in both, to develop an integrated, multi-hazard system. So far, it has been difficult to build even single-use systems except at a national level. A fully operational multi-hazard observing system will require unprecedented cooperation among a wide community of experts and stakeholders. But it would also dramatically improve cost-effectiveness, by both reducing the initial investment and spreading the burden of long-term costs.

The final and most difficult challenge will be to tailor the system to local cultural, social and economic conditions. Although the tsunami warning system must work on a global scale, its users will be local. As with so many things, we need to be thinking globally and acting locally. Civil populations cannot

be educated or warned without accounting for — and benefiting from — local knowledge and concerns. Outreach, education and public awareness efforts will only work if they are woven into national, cultural and local environmental fabrics. For example, in Aceh, Indonesia, it has been suggested that rapid delivery of warnings could exploit the wide distribution of Islamic mosques with loudspeaker systems used for calls to prayer.

Ultimately, the development of the scientific and technical backbone of a tsunami warning system is a global responsibility, but preparedness remains a task for individual nations, or regions. This is the hardest of the three challenges and will require novel mechanisms for cooperation between scientists and social scientists, and between different organizations at the international, national and regional levels.

In particular, the international scientific community must not get carried away with the tantalizing but flawed idea that there is a quick technological fix to these complex societal issues. Instead, we need to broker a process through which countries of any given region come to recognize themselves as the true owners of the system. In their eagerness to help, states or organizations from outside the region might even obstruct the process by which Indian Ocean rim countries come together to plan, create and implement a system. But such a process should develop a true sense of ownership and responsibility. The majority of the lives lost were Asian, and the

countries of that region must be at the forefront of plans to protect themselves in the future.

From 3 to 8 March 2005, UNESCO is hosting the first of two technical meetings intended to foster the development of a tsunami warning and mitigation system for the Indian

Ocean. All of the nations in the region are invited and, along with other interested nations and international organizations, will work together to design a comprehensive work plan and timetable.

The challenge facing these countries, together with the IOC and our global partners, is a substantial one. But unlike so many visionary projects mooted by bureaucrats, the task is both clearly defined and eminently achievable. Let us hope that we are now taking the first step to ensure that the next tsunami — wherever and whenever it inevitably occurs — will not go down in history as a catastrophe, but as a tribute to the ability of science and technology to serve society. ■

Keith Alverson is at the Global Ocean Observing System of the Intergovernmental Oceanographic Commission of UNESCO, 1 Rue Miollis, 75732 Paris, CEDEX 15, France.

Acknowledgements

I thank Thorkild Aarup, Bernardo Aliaga, Patricio Bernal, Ehrlich Desa, Albert Fischer, Paul Mason, Peter Pissiersens and Francois Schindele for contributing many useful thoughts to this article.

Primate viewing

Chimpanzee behaviour shows remarkable regional variation.

The Cultured Chimpanzee: Reflections on Cultural Primatology

by William McGrew

Cambridge University Press: 2004. 244 pp.
£50, \$90 (hbk); £19.99, \$29.99 (pbk)

Tetsuro Matsuzawa

Thirty years ago, two young researchers, William McGrew and Caroline Tutin, visited the Mahale Mountains in western Tanzania. Until then they had been studying wild chimpanzees at Gombe, about 120 km to the north, but on their first day at Mahale they saw two chimpanzees perform a striking behaviour that was completely new to them. The two chimpanzees were sat on the ground facing one another, and were engaged in mutual grooming. At one point, each fully extended one arm overhead and clasped the other's hand. This created a sort of 'A-frame' postural configuration that revealed the armpit of the raised limb, which was then groomed by the other's opposite hand. The two chimpanzees were in perfect symmetry.

Returning to camp the same evening, McGrew and Tutin mentioned their discovery of the 'grooming hand-clasp' to their host, Junichiro Itani. But Itani was unimpressed: "Don't all chimpanzees do this?" he asked. This was a turning point for McGrew, who at that moment realized that scientists had hitherto been labouring under a simple misapprehension: that chimpanzee social life was the same everywhere.

McGrew and Tutin's pioneering report on the evidence for a 'social custom' in wild chimpanzees was published in 1978, but its importance was not fully recognized until later. At about the same time, fieldworkers on the other side of the continent — at Tai in Côte d'Ivoire and Bossou in Guinea — were making observations about the use of stones by West African chimpanzees. At these sites, wild chimpanzees were using stones to crack open hard-shelled nuts containing edible kernels. In contrast, chimpanzees at Gombe were known to eat the mesocarp, flower, pith, resin and cambium of the oil palm but discard its hard-shelled nut — they lacked the elementary stone technology of their West African relatives.

As increasing numbers of papers were published, the behavioural diversity of chimpanzees in the wild became clearer. McGrew's influential *Chimpanzee Material Culture* (Cambridge University Press, 1992) was the first book to paint a clear picture of patterns of culture. It showed that different communities of wild chimpanzees have



Variety show: chimpanzees at Mahale in Tanzania perform the 'grooming hand-clasp' but those at Gombe do not.

different tools and skills, and that not all of this regional variation can be explained by the demands of the physical and biotic environments in which they live.

The Cultured Chimpanzee is a worthy follow-up, introducing a new discipline called 'cultural primatology'. Its emergence came about as a natural extension of our expanding knowledge of cultural differences among wild chimpanzee communities. The book reviews cultural phenomena in other primate species, as well as non-primates such as fish, birds, mammals and cetaceans. According to McGrew, cultural primatology has a cross-disciplinary nature, having aspects of at least four traditional academic disciplines: anthropology, archaeology, psychology and zoology. Do non-human animals have culture? It depends on the definition. Each discipline asks different questions about culture, and uses different methods to answer them. *The Cultured Chimpanzee* has 196 pages of text, but contains 469 references — an indication of his dedication to synthesizing the different approaches, covering all the relevant papers about culture in non-human animals, especially chimpanzees.

When trying to sum up the book, three words spring to mind: clear, simple and deep. As McGrew confesses, he is a naturalist

(as opposed to an experimentalist), but he pays attention to important issues such as imitation and teaching that have been examined in detail in the lab. He may be an empiricist (and not a theoretician), but he creates a unique framework for drawing scattered data together, thereby clarifying what is known and what is not yet known. His logic and his trains of thought are extremely clear. The text is simple to follow, even for non-English readers, and yet the messages are stimulating, heuristic and reach deep into the heart of the matter. In particular, the chapter entitled "Lessons from cultural primatology" will provide young scientists — future protagonists in the development of this new discipline — with plenty of good advice.

McGrew's *Chimpanzee Material Culture* is already recognized as one of primatology's classic textbooks. This 2004 follow-up should receive similarly wide attention and become another milestone in the study of the evolutionary basis of human culture. However, I would, at some future date, like to see a

third book as well, written by the same author on the same topic. As McGrew mentions in the preface, *The Cultured Chimpanzee* was written just before his first visit to Bossou, Guinea, where a small group of 19 chimpanzees uses stones to crack nuts. As a naturalist and empiricist, coming face-to-face with this behaviour has hopefully provided McGrew with material for new and stimulating insights.

Just as McGrew concludes the book by drawing attention to conservation efforts, I would like to conclude this review by stressing the importance and urgency of protecting the chimpanzees and the forests of Africa. Chimpanzees probably once spanned most of equatorial Africa, including at least 25 countries. They probably numbered more than a million just 100 years ago. Today they occur in 22 countries, and an estimate from the World Conservation Union (IUCN) in 2003 put their numbers in Africa at between 172,700 and 299,700. This sudden decrease is linked to various human activities, such as deforestation, poaching and trading in bushmeat, as well as the transmission of diseases. For example, the Bossou community lost 5 of its 19 members to a contagious respiratory disease at the end of 2003. Similar stories are taking place all over Africa. Truly intense

efforts are necessary on our part to prevent the extinction of the cultural variation among chimpanzee communities that we have so recently begun to uncover. ■

Tetsuro Matsuzawa is at the Primate Research Institute, Kyoto University, Kanrin, Inuyama-city Aichi 484-8506, Japan.

Science in court

Laws of Men and Laws of Nature: The History of Scientific Expert Testimony in England and America
by Tal Golan

*Harvard University Press: 2004. 336 pp.
\$49.95, £32.95, €46.10*

Sheila Jasanoff

Change one word — write ‘trials’ in place of ‘laws’ — and this appealingly readable book would just as appropriately be titled *Trials of Men and Trials of Nature*. For trials are the stuff of Tal Golan’s engaging narrative as he briskly guides his readers through some of the formative moments in a century or so of scientific expert testimony in English and American common law. Men’s wits and character are on trial throughout these cases, as experts from varied fields vie to position themselves, their skills and their specialist knowledge at the service of the courts. Nature, too, is often on trial, for the outcomes in the cases that Golan skilfully dissects usually turn on who is right about the way the world works, whether in explaining the silting up of a harbour on the North Sea coast of Norfolk, distinguishing human from animal blood, displaying an X-ray picture of a badly set bone, or diagnosing, through bodily measurements, the likely truthfulness of a witness’s testimony in a murder trial.

On one level, Golan’s well-chosen selections from the annals of nineteenth-century litigation confirm contemporary prejudices about the relationship between science and law. As in most bad marriages, encounters between the two professions seem unavoidable and yet are sources of profound miscommunication. Since the early days of the Industrial Revolution, there has been no issue so arcane, nor claim so untenable, that an expert cannot be found to help defend it in court. Experts are available for hire in cases ranging from disputes over land use and environmental degradation to criminal identification, medical malpractice and the insanity defence.

But more knowledge does not necessarily mean more illumination. The common law’s adversarial genius can divide entire communities of knowledge-holders into opposing camps, each seemingly more interested in its side winning than in arriving at the truth. The notion of science as a disinterested fact-finding practice flies out of the courtroom

window as floods of expert testimony frustrate judges, confuse juries and make trial outcomes increasingly unpredictable. With experts dominating litigation, laws — natural or human-made — seem less and less relevant to dispensing justice.

Golan on the whole shares the sense of deepening crisis that has gripped Anglo-American courts since the advent of professional expert testimony. “Alas,” he repeatedly exclaims, as he recounts one story after another in which unresolvable battles between partisan experts took over from any impartial attempt to discern the facts of the case. As a historian, Golan is not primarily concerned with solutions, but he does not hide his yearning for a more orderly process in which judges would proactively scrutinize, and perhaps exclude, expert claims, while juries would be selected on the grounds of technical competence rather than generic civic capacity. In this spirit, he approves of the US Supreme Court’s 1993 decision in the case of *Daubert v. Merrell Dow Pharmaceuticals*, a lawsuit over birth defects allegedly caused by the drug Bendectin, which roundly affirmed the duty of federal judges to act as gatekeepers in relation to expert testimony. Judges, the *Daubert* case declared, should screen expert evidence in accordance with scientists’ criteria for determining whether proffered testimony is based on reliable theories and methods. Judges, in short, were asked to think like scientists — and, in so doing, to keep unreliable evidence away from overly credulous juries.

Unfortunately, as Golan’s book demonstrates, the problem of expert testimony is too complicated to be solved through the simple expedient of substituting judges for juries. The historical cases impressively reconfirm a point often made by scholars of science studies: the science that courts need, along with the methods for generating it, frequently evolves under the prod of litigation, as part and parcel of the adversary process. Courts in technology-intensive societies are as much agents for producing new knowledge as sites for applying what is already known. Facts are generated, often under severe material and temporal constraints, to fill in gaps in available knowledge and to address the uncertainties of actual cases. Judges, then, may go seriously astray in excluding such evidence because it does not meet the standards of established science. Indeed, because *Daubert*-like challenges tend to favour corporate defendants more than plaintiffs, exclusion-minded judges may turn out in practice to be thinking more like corporations than like disinterested scientists.

What will happen if the law’s dependence on expertise intensifies still further? Golan hopes that salvation will come from within the legal system — and in an unexpected way his wish may be granted. The trial itself may gradually yield to methods of dispute

resolution that turn less on the theatrics of the adversary process. Litigation statistics in the United States suggest that trials are becoming a thing of the past, a consequence no doubt of spiralling costs, of which expert testimony is a not inconsiderable fraction. But are backroom bargains, out-of-court settlements and sealed court records desirable substitutes for litigation? This is a question that those committed to both truth and justice may reasonably ask. ■

Sheila Jasanoff is professor of science and technology studies, John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA.

Sizing up the world

Measurement Theory and Practice: The World Through Quantification

by David J. Hand

Hodder Arnold: 2004. 320pp. £45, \$60

Stephen Senn

Measurement theory provides a similar touchstone in science to linguistic theory in philosophy. Some see it as fundamental, others as trivial. Most scientists regard it as a distraction, as they seek to theorize and measure, but not to theorize about measurement. It is surprising how many statisticians are largely indifferent to the nature and purpose of measurement. To be sure, there are many statistical theories of errors in measurement, and plenty about probability, but these are not the same as theories about measurement itself. Statisticians have



A measured approach: France adopted the metric system in the late eighteenth-century.

Museum

A medical history

Whether your interest lies in anatomy, pathology, surgery or the history of science, or, like myself, you are just curious, the reopening last month of the Hunterian Museum at the Royal College of Surgeons of England in London is an event to be celebrated.

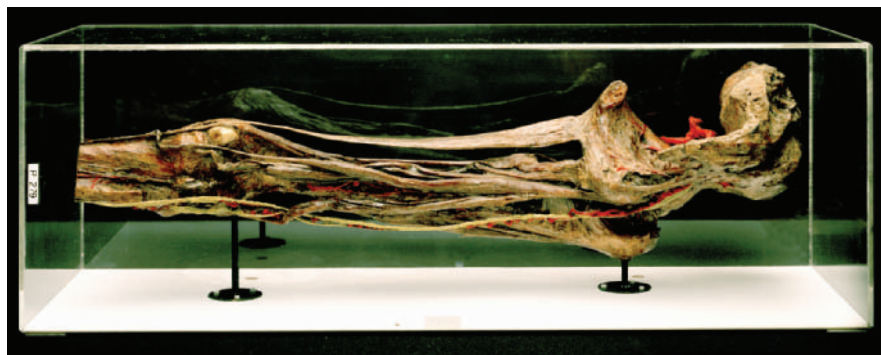
Although only 3,500 of the Hunterian's collection of some 65,000 items survived the bombing of the museum in 1941 (right), the remainder of this great eighteenth-century 'cabinet of curios' still ranges over an extraordinary and wonderful breadth of natural science. From the treatment of gunshot wounds and syphilis, through the anatomy of the extinct giant deer, to tooth transplants, the life cycle of bees and a dissection of a coachman's leg (below right), there is something for everyone in this amazing collection, which has undergone a £3.2-million (US\$6-million) refurbishment.

The museum's founder, John Hunter (1728–93), was one of three brothers from Scotland who sought fame and fortune as surgeons in London. William, the eldest, was already well established when John joined him as an apprentice in 1748. After a career as an army surgeon, John married the talented Anne Home, who established a fashionable salon that attracted the great and the good of the day, from James Boswell to Joseph Banks. John was a workaholic who became successful as a surgeon, pathologist, teacher and experimentalist, and helped to lay the foundations of modern medicine. This is suitably celebrated by the excellent new displays put together by curator Simon Chaplin and his team.

William Hunter's collection is at the Hunterian Museum in Glasgow.

Douglas Palmer

♦ www.rcseng.ac.uk/services/museums



THE ROYAL COLLEGE OF SURGEONS OF ENGLAND

HUNTERIAN MUSEUM

a tendency to limit their contributions in research collaborations to advising or determining how measurements should be analysed, and how many should be taken, rather than what the measurements should be. Even in the field of experiment design, the emphasis is on choosing patterns of inputs to the experiment, rather than advising on the measurement of outputs.

My field is medical statistics and although I take some measurement issues seriously, I shy away from others. I always find, for example, that whenever I have anything to do with quality of life, mine suffers and I avoid thinking about it accordingly. An honourable exception among statisticians is David Hand, who originally trained as a mathematical physicist. Since his conversion to being a statistician he has worked in psychology and medicine, as well as on economic and financial topics such as credit scoring. Hand, who is professor of statistics at Imperial College London, has been worrying about issues concerning measurement for years and has now written this thought-provoking monograph.

There are two different sorts of chapter in the book. The first four cover matters of fundamental importance for all sciences: a general introduction is followed by discussions of the nature and process of measurement and accuracy. The last five chapters cover various sciences — psychology, medicine, the physical sciences, economics and the social sciences — and include a final remain-

der chapter to sweep up all terms not covered by the preceding expansion. My one major criticism is that there isn't a chapter on statistics itself: there are just three pages on probability in the final chapter. More could have been said about this. For example, is measure theory fundamental or a fundamental mistake as maintained by Glen Shafer and Vladimir Vovk in their book *Probability and Finance* (Wiley, 2001)? Other omissions include any discussion of the measurement of political preferences and, for example, Condorcet's paradox and Arrows' impossibility theorem and their implications for the impossibility of perfect voting systems.

However, the book is full of wonderful things. Here is Hand writing about Luce's principle governing the classes of possible relationships between variables, which has implications as to the sort of scientific laws that are possible: "To me, when I first heard of the idea and saw its implications, it seemed remarkable, lending extraordinary power to the search for scientific laws." I have also now had such a revelation, but its source was Hand's book and the valuable discussion of Luce that it contains, both in chapter 2 (whence the quote) and later in chapter 7 on the physical sciences. This latter chapter also has excellent discussions on dimensional analysis and the implications for regression coefficients.

Hand also sheds light on the baffling and notorious 'two-envelopes puzzle' (also

known as the 'exchange paradox'). You are given a choice of two envelopes and reliably informed that one contains twice as much money as the other. Having picked one, but not yet opened it, you argue: "If I exchange, I double my money with a probability of one-half, and halve it with probability of one-half, and since half of two plus half of a half is one-and-a-quarter, my expectation is greater if I exchange." Having exchanged, you can then use the same argument to change back again.

Many other matters are expertly touched on too. To pick some, not at all at random, I found the discussions of the various forms of indirect scaling, of psychophysics, of indices in economics, and on league tables particularly interesting, and will find much of what is in the chapters on medicine and psychology useful in my work. The book is also pleasantly sprinkled with historical observations, interesting quotations and anecdotes. For instance, we learn of Claude Litre, born in Margaux in the heart of the Medoc (whose name speaks volumes, but of fiction in this case).

This book ought to be on every statistician's shelves and on those of many other scientists as well. The author concludes that "measurement is what distinguishes the civilized from the uncivilized". He is to be congratulated for this stimulating contribution to civilizing his fellow scientists. ■

Stephen Senn is in the Department of Statistics, University of Glasgow, Glasgow G12 8QQ, UK.

Schrödinger's mousetrap

Part 7: Lessons from the past.

Nicola Spaldin

"How could he have known?" Ludmilla Shlomiuka asked herself frantically. "How could he possibly have known?"

The interview with Inspector Lister had started uneventfully. She had expected the question about her relationship with Rufus Jaeger, of course. The stereotype of the beautiful young researcher and the brilliant professor just lends itself to gossip. And it's no secret that de Bruijn does his part to propagate the rumours. It's sad that he's so bitter and confused. He blames Jaeger for his lack of professional advancement — now that is just plain silly — and he envies the excellent collaboration she has with him. Yes, he's been obsessed with her since the day she joined the research group! Lister had asked whether she had received any strange messages from him recently. Of course, she said no. At least no more strange than his usual attentions.

And where was she during the morning break? In the hallway, where the coffee was served, talking to various friends and colleagues. Yes, she also talked to Jaeger; in fact they were discussing the demonstration when de Bruijn came along looking for an argument. Everyone must have noticed him squabbling with Jaeger, it was really quite awkward.

She had also carefully prepared her response to cover her husband: "Dmitri! He suffers from Parkinson's disease and is very sick. Thank goodness that we have such good medical facilities in this country! This morning he was heavily drugged and could barely walk." Surely Lister did not think that Dmitri had been jealous and... She'd pulled that one off rather well, she thought, complete with some convincing tears and sobs. Lister had been quite embarrassed.

And yes, of course she felt sorry for Feng. It was a pity that he had not calibrated the detector correctly, but at least his research was back on track now. No, she hadn't been an author on the retracted paper; she was working on a different project at the time. Yes, she'd heard a little about his latest project with



Petra Pruszczyński; it was apparently quite a breakthrough in optics. She was very much looking forward to reading the paper; there were rumours that *Nature* was rushing it through for publication. No, she didn't think that he was the murdering type.

But her mother? How could he have known? It had been almost 35 years since the state department sent its science envoy to Dubna! She only knew herself because of the old handwritten lab books in her mother's attic. There would be official records listing Jaeger as a member of the envoy of course, but how could Lister have seen them? Surely they were buried in a dusty cabinet somewhere in Washington. There were no leads in the open literature; joint publication would never have been allowed, and anyway there had not been a collaborative project to write about. Just the lab books.

How enthralled she had been when she had found them! Everything was there. The fundamental mathematical development and suggestions for implementation that were decades ahead of their time. The ideas on which Jaeger had built his career.

Growing up, she had heard her mother talk occasionally about the visit. It had been so exciting for the Soviet researchers to discuss their work with outsiders. The nuclear programme was top secret, of course, but they had been

given permission to talk about their basic research, maybe even encouraged to show off a little. At the time her mother had been developing some theoretical aspects of what we now call quantum entanglement. There were very few hours to devote to it because of her commitments to the nuclear programme, but she was playing with some new ideas more-or-less in her spare time. In fact, she wasn't even sure whether they were new or not, as her access to the literature was restricted, but it was interesting to develop them anyway.

Her mother had had such a passion for fundamental physics.

It's funny the things that people remember. Years later she had mentioned that they were separated from the visitors at lunch, not to prevent social interaction, but because the foreigners were served a better quality meal!

And they had been allowed to turn on the heating in the lab, even though it was only October. And she had recounted with great fondness her discussions with the young American who had been so courteous and showed such interest in her results...

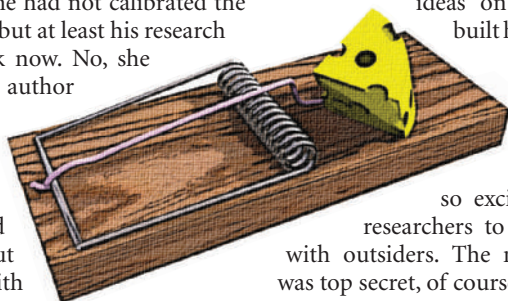
Shlomiuka tried to compose herself for the rest of the interview. Why had she applied to join Jaeger's group? Because he was indisputably the leader in the field. No, no other reason. Well, yes, the healthcare for Dmitri was a bonus of course. And he had selected her because she had produced an excellent PhD thesis and she was very good at her work. Yes, she was sure that was the only factor. He probably didn't even recognize the name Shlomiuka. Of course she was a little curious, but no she wasn't resentful. Yes, he was an excellent mentor to her; he had been promoting her for a faculty position of her own before...

Just as she started to collect herself, Lister pulled out his trump card: "And finally, Dr Shlomiuka, is it a coincidence that the state department envoy was sent to Dubna 35 years ago, and that you just celebrated your 34th birthday?"

To be continued...

Nicola Spaldin is in the Materials Department, University of California, Santa Barbara, Santa Barbara, California 93106-5050, USA.

Who do you think killed Rufus Jaeger? Catch up on all the evidence and vote for your suspect at www.nature.com/news/mousetrap



Coupling and cross-presentation

William R. Heath and Francis R. Carbone

Studies of cultured cells have revealed how the immune system may use intercellular pores to convey information that is important in initiating antiviral responses and in limiting the spread of infections.

Our immune system has an ingenious way of dealing with microorganisms that invade our cells. Each cell displays tiny protein fragments (peptides) on its surface, representing most of the proteins that are found within. If the cell is infected by, for instance, a virus, then the evidence is likewise displayed. The immune system's killer T cells screen the displayed peptides; if they detect a viral peptide, they destroy the infected cell, and the virus with it.

The source of peptides presented by a cell was thought to be limited to those proteins synthesized within that cell. On page 83 of this issue¹, however, Neijssen *et al.* show that peptides can be transferred between cellular neighbours, through small molecular pores called gap junctions. This offers a mechanism for extending the destruction of infected tissue to a few surrounding cells, thus efficiently limiting virus spread. It also provides a way of supplying viral peptides to dendritic cells, which kick-start immunity.

To be displayed on the cell surface, peptides must be slotted into receptors known as major histocompatibility complex class I (MHC I) molecules, and it is these peptide-receptor complexes that are screened by killer T cells. The presentation of peptides — also known as antigens — provides the T cells with a non-invasive method of looking inside target cells and determining whether or not they are harbouring infectious agents (or mutated, potentially cancerous, proteins). For this to work, it is important that each screened cell presents only peptides derived from its interior (endogenous peptides), and not those that it picks up from the surrounding milieu (exogenous peptides). Otherwise, healthy bystander cells that are simply bathed in protein debris from infected cells might be destroyed unnecessarily.

This requirement presents a problem, however. Killer T cells cannot start an immune response on their own; for this, they require dendritic cells, which, by virtue of their expression of specialized antigen-presenting machinery, co-stimulatory molecules and inflammatory (cytokine) proteins, are able to present antigens to naive T cells to initiate immunity². But if the access of exogenous proteins to the MHC I pathway is strictly prevented, how can immunity be initiated in response to viruses that simply do not infect dendritic cells — or to viruses that infect these cells but can inhibit the display of

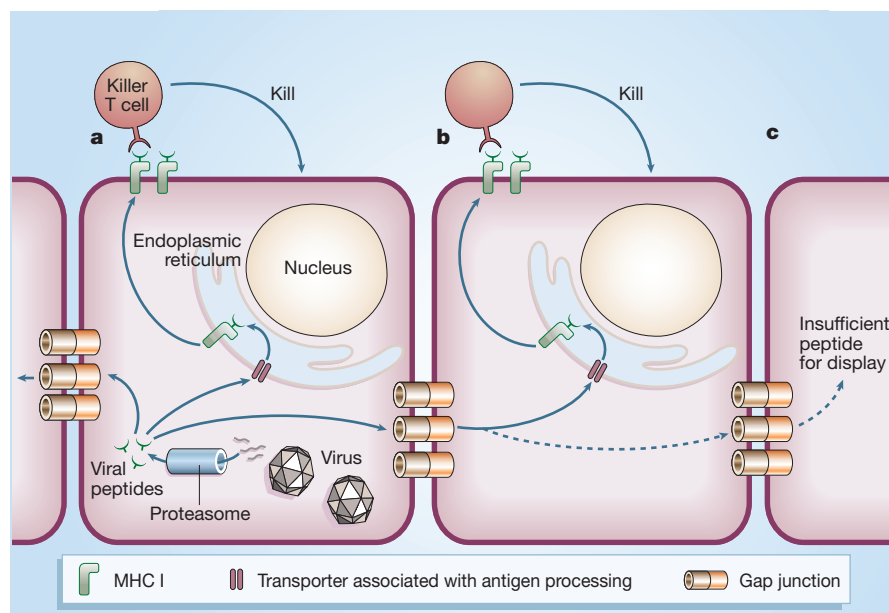


Figure 1 'Cross-presentation' by peptide transfer through gap junctions. a, Peptides derived from a cell's own proteins are presented on the cell surface (not shown). If the cell is infected by a virus, viral peptides will likewise be presented. The viral proteins are first chopped up in the proteasome; resulting peptides are transported into the endoplasmic reticulum, where they are paired with major histocompatibility complex I (MHC I) molecules and then moved to the cell surface. Killer T cells detect viral peptides and destroy the infected cell. b, Neijssen *et al.*¹ find that peptides up to ten amino acids long (derived from viral or cellular proteins) may also be transferred from one cell to its neighbour through gap junctions. These peptides follow the classical MHC I pathway and are displayed on the surface for recognition by killer T cells. c, Cells farther away probably do not receive enough viral peptides to be displayed.

endogenous peptides on MHC I molecules?

Some resolution of this issue was provided by Bevan's discovery³ that specialized immune cells, including dendritic cells, can capture proteins from other cells (potentially virus-infected ones) and direct them into their own MHC I pathway. This unusual ability to transfer exogenous proteins into the MHC I pathway has been referred to as cross-presentation — primarily because proteins cross from one cell to another. The mechanistic basis for cross-presentation is not well understood, but several pathways have been described^{4,5}.

Neijssen *et al.*¹ have now discovered another such pathway, by which cells obtain peptides from their neighbours through gap junctions (Fig. 1). These junctions are small molecular pores formed between cells by connexin proteins, and are generally used for the intercellular transport of nutrients and other small molecules⁶. Neijssen *et al.* show that connexin 43 can form channels that

allow peptides up to ten amino acids long to pass between cells. Although these authors do not directly examine transfer to dendritic cells, they do show that a close cellular relative, the monocyte, can use coupling through gap junctions to load its own MHC I molecules with viral peptides derived from adjacent, infected cells.

Neijssen *et al.* also show that Langerhans cells — one of several subtypes of dendritic cell⁵ — form gap junctions with surrounding skin cells (keratinocytes). This provides the Langerhans cells with a potential method of sampling keratinocyte peptides before migrating to the lymph node, where they initiate immunity to skin infections. Consistent with this idea is the report that Langerhans cells can cross-present peptides expressed by keratinocytes⁷. And, interestingly, viruses such as herpes simplex virus (HSV) can block gap-junction transport⁸, perhaps explaining why Langerhans cells fail to cross-present HSV peptides during skin infection⁵.

Fortunately, however, other types of dendritic cells can participate in inducing immunity to HSV, thus circumventing this blockade — perhaps by using other mechanisms of cross-presentation. For example, various investigations have shown that cross-presentation can involve the transfer of larger protein fragments⁹. These would not pass through the small pores formed by gap junctions, emphasizing that gap-junction-mediated cross-presentation is not the only mechanism for antigen transfer.

As well as providing another possible mechanism for initiating immunity by dendritic cells, the gap-junction-mediated cross-presentation described by Neijssen *et al.* offers an interesting method of efficiently limiting the spread of replicating virus. The authors show that not only will a cell expressing viral proteins be killed by T cells, but so will its closest neighbours — because they present viral peptides obtained through gap junctions. Extending the destruction to adjacent cells may provide a 'fire-break' around an infection, ensuring that if low levels of virus have spread to surrounding cells, but have yet to produce sufficient protein to allow recognition, such cells will still be eliminated. The rapid degradation of peptides within the cell's cytosol means that the spread of peptides through gap junctions will be rather

limited, probably allowing the targeting of adjacent cells but not those more than one cell distant from the infection. Thus, the integrity of targeting should be maintained, with only limited bystander destruction.

At present, the physiological role of cross-presentation in immunity to viruses and tumours is unclear^{10,11}. But a better understanding of the mechanisms by which it can occur — as provided by the work Neijssen *et al.*¹, for instance — should lead to the resolution of this important issue. ■

William R. Heath is at the Walter and Eliza Hall Institute, 1G Royal Parade, Parkville, Victoria 3050, Australia.

e-mail: heath@wehi.edu.au

Francis R. Carbone is in the Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria 3010, Australia.

e-mail: fcarbone@unimelb.edu.au

1. Neijssen, J. *et al.* *Nature* **434**, 83–88 (2005).
2. Banchereau, J. & Steinman, R. M. *Nature* **392**, 245–252 (1998).
3. Bevan, M. J. *J. Immunol.* **117**, 2233–2238 (1976).
4. Yewdell, J. W., Norbury, C. C. & Bennink, J. R. *Adv. Immunol.* **73**, 1–77 (1999).
5. Heath, W. R. *et al.* *Immunol. Rev.* **199**, 9–26 (2004).
6. Oviedo-Orta, E. & Evans, W. H. *J. Leukoc. Biol.* **72**, 636–642 (2002).
7. Mayerova, D., Parke, E. A., Bursch, L. S., Odumade, O. A. & Hogquist, K. A. *Immunity* **21**, 391–400 (2004).
8. Fischer, N. O., Mbuy, G. N. & Woodruff, R. I. *J. Virol. Methods* **91**, 157–166 (2001).
9. Ploegh, H. L. *Science* **304**, 1262–1263 (2004).
10. Melief, C. J. *Eur. J. Immunol.* **33**, 2645–2654 (2003).
11. Zinkernagel, R. M. *Eur. J. Immunol.* **32**, 2385–2392 (2002).

Astronomy

Blasts from the radio heavens

S. R. Kulkarni and E. Sterl Phinney

There is no coherent explanation for newly observed salvos of radio waves emanating from a direction near the Galactic Centre. Are they from a new type of stellar object? The search is on for similar radio emitters.

For thousands of years, we self-important humans have interpreted transient heavenly events as omens. The Chinese emperor was the Son of Heaven and paid a retinue of astronomers to keep careful track of comets and other 'guest stars' (novae and supernovae) to predict earthly catastrophes. Modern astronomers are mostly paid for different reasons, yet continue to discover new kinds of transients, which have delivered handsome dividends in our understanding of stellar death and corpses (white dwarfs, neutron stars and black holes). The terms supernova, nova, X-ray transient, γ -ray burst and magnetar have crept into the lexicon of most readers of *Nature*.

On page 50 of this issue, Scott Hyman *et al.*¹ report a bright bursting radio source near (in projection at least) the centre of our Galaxy. They suggest that the object (dubbed GCRT J1745–3009) is a prototype of a new class of particularly bright, coherently emitting radio transients. Because the distance and precise position of the source are as yet

unknown, more mundane explanations are still possible. But the manner of its discovery, and the potentially exciting interpretation, will inspire more dedicated searches for radio transients.

Now the essential facts. While observing the central region of our Galaxy at radio wavelengths, Hyman and colleagues discovered five strong bursts, each lasting about 10 minutes, and separated by about 77 minutes, coming from the same 10-arcsecond region of the sky. No emission, steady or otherwise, is seen in subsequent (and archival) searches or between the bursts.

The source is not well enough localized to identify counterparts at other wavelengths, so its distance is unknown. GCRT J1745–3009 could, like most stars in its direction, be near the centre of our Galaxy (about 24,000 light years or 8,000 parsecs distant), in which case its radio luminosity is around an impressive one-hundredth that of the Sun. But because the centre of our Galaxy is so interesting, astronomers tend to

stare there more than anywhere else. So it is possible that the source is much nearer (say, 300 light years or 100 parsecs), less luminous, and only coincidentally projected near the Galactic Centre.

The duration of the burst limits the size of the source to less than the distance travelled by light over the burst duration. Armed with this knowledge, we can compute the equivalent 'black-body' temperature of the emitter. Cosmic radio sources have a natural thermostat² that normally restricts this brightness temperature to less than 10^{12} kelvin. But the brightness temperature of GCRT J1745–3009, the radio source observed by Hyman *et al.*, exceeds this if it does indeed lie farther from Earth than 100 parsecs. In some cases, temperatures higher than the 10^{12} -kelvin thermostat are seen through a form of trickery³ involving special relativity: when emitting matter is racing towards Earth at nearly the speed of light, much higher apparent brightness temperatures will be inferred.

Galactic examples of such astronomical tricksters are black holes⁴ and neutron stars⁵ in binary systems accreting mass from a companion star. GCRT J1745–3009 could be one of those objects. However, the known examples have prominent X-ray emission, whereas no X-ray emission from GCRT J1745–3009 has been reported in other studies (the RXTE, ROSAT and ASCA space missions). On the other hand, if the roughly 77-minute interval between the source's radio bursts is an orbital period in an accreting binary system, only the smallest star or a white dwarf can fit in the tiny orbit. The accretion rate in such a binary system would be low, and the accretion might also be radiatively inefficient^{6,7}, so it could hide well below the X-ray limits. It is possible that the radio source, modulated by absorption of the radio waves in a stellar wind, is an 'X-ray quiet, radio-loud' X-ray binary, similar to certain types of active galactic nuclei⁸, but with stellar mass.

Instead, the radio source could also genuinely beat the 10^{12} -kelvin thermostat by emitting its radiation coherently. Coherent emission requires organized electrons. Terrestrial examples include radio stations, masers and lasers. Coherent emission when seen in heavenly objects is always a source of wonder. Our Solar System has some coherent emitters: cyclotron masers in planetary aurorae, long-wavelength radio emission from the Io–Jupiter system, and solar flares. However, the impressive luminosity of GCRT J1745–3009, even if it is as near as 100 parsecs, is not compatible with such systems. Outside the Solar System, we know of only three types of coherent emitter: masers (ruled out for GCRT J1745–3009 because it does not have the appropriate narrow bandwidth), magnetically active 'flare' stars, and — most spectacularly — pulsars.

Radio emission from flare stars is usually (but not always) highly polarized⁸. The dwarf

star or brown dwarf would also be seen at optical wavelengths. There are potential candidates within the uncertain position of GCRT J1745–3009, so this remains a possible, though unlikely, explanation.

Could GCRT J1745–3009 be a pulsar? Hyman *et al.* note that the five bursts they record appear in rapid succession with a period of about 77 minutes. The rotational energy lost by a normal neutron star (or white dwarf) rotating this slowly would be inadequate to power the observed radio bursts. Thus, by a process of elimination, Hyman and colleagues argue that they have uncovered a new class of coherent emitter.

In our opinion, the claim of a new class is plausible but not beyond doubt. As discussed above, the bursts could still be incoherent emission from an accreting binary star with a whittled-down companion and a relativistic jet but suppressed X-ray emission. If the source turns out to be nearer than the Galactic Centre, it could be one of several previously known types of coherent radio source, including an isolated or binary flaring (brown) dwarf star or magnetized white dwarf, or a nulling radio pulsar (a pulsar that broadcasts pulses only sporadically).

This last seems to us to be the most plausible conventional alternative. PSR 0826–34, for example, is a pulsar that can shut itself off for periods ranging from tens of minutes to eight hours⁹. PSR J1752 + 2359 is characterized by 45-second bursts of emission that appear roughly every five minutes¹⁰, like GCRT J1745–3009 but speeded up by an order of magnitude.

GCRT J1745–3009 will cause a stampede of further observations: searches for pulsations and quiescent emission in radio, infrared and X-ray bands. But perhaps even more important is the possibility that the radio heavens contain other fast radio transients (which, in anticipation of a trove of discoveries, we nickname 'burpers'). Sensitive radio telescopes and arrays currently lack large fields of view. Fortunately, the construction of several new radio facilities with wider fields of view are being contemplated, and one is already funded¹¹. Radio astronomy is poised to deliver new bursts of excitement. ■

S. R. Kulkarni and E. Sterl Phinney are in the Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, California 91125, USA.

e-mails: srk@astro.caltech.edu; esp@tapir.caltech.edu

- Hyman, S. D. *et al.* *Nature* **434**, 50–52 (2005).
- Readhead, A. C. S. *Astrophys. J.* **426**, 51–59 (1994).
- Rees, M. J. *Nature* **211**, 468–470 (1966).
- Mirabel, I. F. & Rodríguez, L. F. *Annu. Rev. Astron. Astrophys.* **37**, 409–443 (1999).
- Fender, R. *et al.* *Nature* **427**, 222–224 (2004).
- Rees, M. J., Phinney, E. S., Begelman, M. C. & Blandford, R. D. *Nature* **295**, 17–21 (1982).
- Esin, A. A., McClintock, J. E. & Narayan, R. *Astrophys. J.* **489**, 865–889 (1997).
- Bastian, T. S. *Sol. Phys.* **130**, 265–294 (1990).
- Durbin, J. M. *Mon. Not. R. Astron. Soc.* **186**, 39P–41P (1979).
- Lewandowski, W. *et al.* *Astrophys. J.* **600**, 905–913 (2004).
- LOFAR www.lofar.org

Evolutionary biology

The hydrogenosome's murky past

Michael W. Gray

The evolution of specialized cellular powerhouses called hydrogenosomes has long confounded biologists. The discovery that in some cases they have their own genome sheds some much-needed light on the issue.

Hydrogenosomes are double-membraned subcellular structures that generate hydrogen while making the energy-storage compound ATP. They are found in certain eukaryotic (nucleus-containing) microbes that inhabit oxygen-deficient environments¹. The evolution of the hydrogenosome has remained obscure, mainly because these organelles seemed not to have a genome^{2,3} — until now. On page 74 of this issue, Boxma *et al.*⁴ report the characterization of what seems to be an authentic hydrogenosomal genome in the anaerobic microbe *Nyctotherus ovalis*, an inhabitant of the termite hindgut.

In eukaryotes that live in oxygen-rich (aerobic) environments, organelles called mitochondria are responsible for making ATP. Although an evolutionary relationship between hydrogenosomes and mitochondria has been postulated, this hypothesis remains contentious^{2,3}. Mitochondria contain a small genome (mtDNA) that retains traces of their evolutionary origin from a bacterial symbiont^{5,6}. Interestingly, the hydrogenosomal DNA isolated by Boxma *et al.* exhibits hallmarks of a bona fide mitochondrial genome.

Adding to this story are two recent

papers^{7,8} that probe the evolutionary history of the hydrogenosome from another anaerobic microbe, the parasite *Trichomonas vaginalis*. The absence of a hydrogenosomal genome in this organism⁹ makes it a challenging task to infer the origin of its hydrogenosome. Indeed, on this point, the two groups^{7,8} come to rather different conclusions, even though they analyse the same *Trichomonas* hydrogenosomal proteins.

In animals and fungi, the mitochondrial genome encodes a small number of essential inner-membrane proteins (components of respiratory complexes I–IV and complex V, a specialized type of ATP-synthesizing enzyme) that function in electron transport and ATP production⁵. In addition, mtDNA specifies the RNA components of the mitochondrial protein-synthesis system and, in plants and many algae and protozoa, some of the proteins of this system too⁵.

The report by Boxma *et al.*⁴ extends their earlier observations¹⁰ — which were provocative but not compelling — that the hydrogenosome of *Nyctotherus* might contain DNA. Having purified *Nyctotherus* hydrogenosomes, the authors isolated a 14-kilobase stretch of DNA and sequenced it⁴. They identified genes that encode

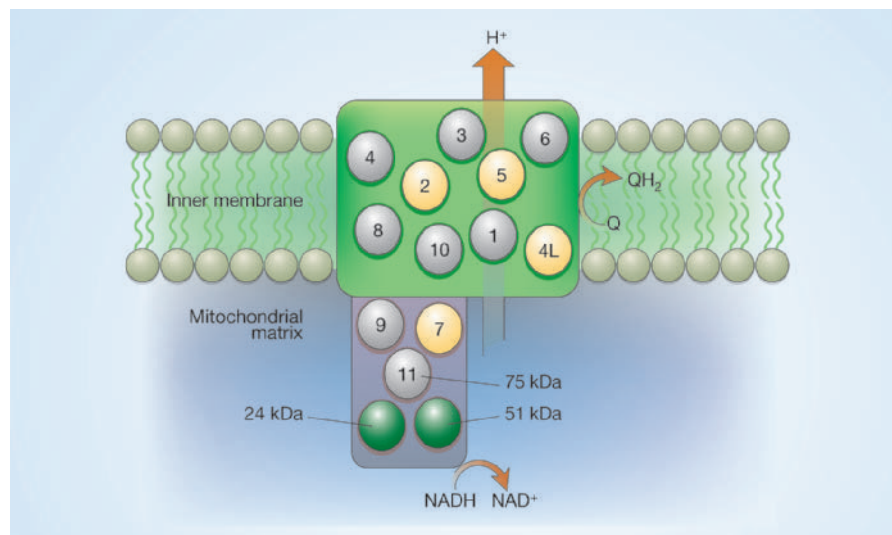


Figure 1 Subunits of mitochondrial respiratory complex I. The membrane-integrated (green rectangle) and peripheral (purple rectangle) regions include numbered subunits that are encoded in one or more mitochondrial genomes; for example, animal mtDNA specifies seven subunits, 1–6 and 4L. Several subunits (yellow) have been identified in *Nyctotherus ovalis* hydrogenosomal DNA⁶. Genes encoding the 51- and 24-kDa subunits (dark green) have only been found in nuclear genomes. A nucleus-encoded 75-kDa subunit has also been reported in *N. ovalis*⁶. In this step of respiration, the oxidation of nicotinamide adenine dinucleotide (NADH) and reduction of ubiquinone (Q) provide protons and electrons to be passed along the respiratory chain, eventually producing ATP and water.

homologues of mitochondrial proteins; that is, the mitochondrial and hydrogenosomal counterparts are close relatives with similar sequences. These genes encode four subunits of complex I (Fig. 1), and two proteins and two RNAs from the protein-synthesis system. The properties of these sequences — for instance, characteristic codon-usage patterns and a similarity to mitochondrial genes from aerobic microbes of the same group as *Nyctotherus* (the ciliate protozoa) — make a convincing case that this DNA is part of an mtDNA-like hydrogenosomal genome.

Additionally, Boxma *et al.*⁴ identify several proteins in *Nyctotherus* that are encoded by genes in the nucleus but are typically transported to and function in mitochondria; these include three additional subunits of complex I (of molecular mass 24 kilodaltons (kDa), 51 kDa and 75 kDa; Fig. 1) and components of complex II. Phylogenetic reconstructions aimed at inferring the evolutionary history of these proteins show an affiliation with mitochondrial (specifically ciliate) homologues. Not unexpectedly, biochemical analyses suggest that *Nyctotherus* hydrogenosomes do not have complexes III and IV, which are responsible for the final stages of aerobic respiration. Nor is there any evidence of a mitochondrial-type ATP synthase (complex V) in this organism.

These and other observations imply that the *Nyctotherus* hydrogenosome represents an intermediate form between mitochondria, which possess a membrane-bound electron-transport chain, and previously characterized hydrogenosomes, which do not — a “true missing link”, in the words of the authors. In parallel, the results suggest that the *Nyctotherus* hydrogenosomal genome, whose total size, shape and gene content have yet to be determined, is probably a reduced ciliate-type mtDNA, lacking those mtDNA-encoded genes that normally specify components required to construct a complete mitochondrial respiratory chain.

The genome-less *Trichomonas* hydrogenosome has been much less forthcoming about its evolution, with sequence-based analysis necessarily limited to nuclear genes that specify the constituent proteins of this organelle. The simultaneous discovery by two groups^{7,8} of *Trichomonas* homologues of the 51- and 24-kDa components of mitochondrial complex I (Fig. 1) is a notable development. These proteins (termed Ndh51 and Ndh24, respectively) are the first *Trichomonas* counterparts of components of the mitochondrial respiratory chain to be identified. However, the two groups differ sharply in their conclusions about the evolutionary origin of these proteins, and hence of the hydrogenosome itself.

Both groups used a standard, rigorous approach for reconstructing evolutionary relationships by comparing protein sequences. However, Hrady *et al.*⁸ conclude

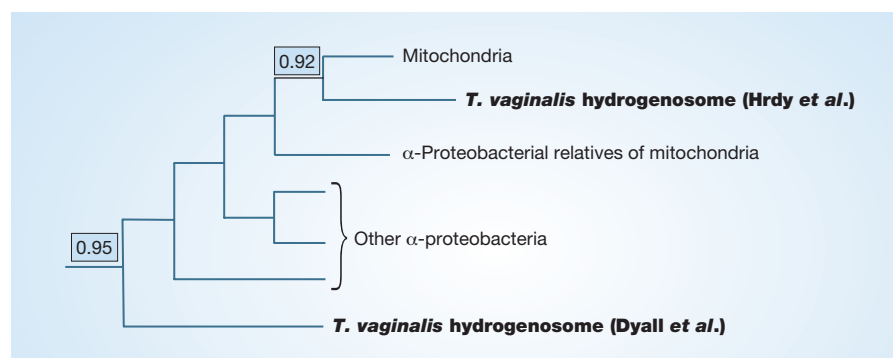


Figure 2 The conflicting evolutionary positions of the *Trichomonas vaginalis* hydrogenosome. In phylogenetic reconstructions based on an alignment of the Ndh51 (51-kDa subunit) protein sequence, Dyall *et al.*⁷ place the *T. vaginalis* hydrogenosome at the base of the α -proteobacterial lineage, not specifically related to mitochondria, whereas Hrady *et al.*⁸ position the hydrogenosome as a specific relative of mitochondria, to the exclusion of α -proteobacteria. Numbers are statistical probabilities that strongly support the associated branches. (Figure courtesy of R. Watkins.)

that *Trichomonas* Ndh51 shares a specific common ancestry with its mitochondrial counterpart, whereas Dyall *et al.*⁷ argue that it does not (Fig. 2). So, why the difference, and who is right?

These conflicting conclusions illustrate a common conundrum in using molecular-sequence data to infer ancient evolutionary events. In parasites such as *Trichomonas*, whose position in the eukaryotic lineage is uncertain to begin with, protein sequences tend to change relatively rapidly in the course of evolution. This can confound their accurate placement in phylogenetic trees, causing so-called long branches. Moreover, *Trichomonas* Ndh51 proved to have a very different amino-acid composition from its counterpart in other organisms, another phenomenon that can severely compromise phylogenetic analysis.

Hrady *et al.*⁸ tried to offset the bias caused by the divergent amino-acid composition by assigning each of the 20 possible amino acids to one of six groups of amino acids that have similar chemistries and commonly replace one another in protein sequences. They then reconstructed the alignment of the Ndh51 and comparison sequences using just the six groups of amino acids and reanalysed the data. This technique has the effect of shortening long branches and homogenizing the amino-acid composition among compared sequences. Using this additional approach, Hrady *et al.* deduced a common origin for the *Trichomonas* and mitochondrial 51-kDa proteins (Fig. 2).

Several points emerge from these three reports. First, Boxma *et al.*⁴ are the first to show that a putative evolutionary relative of the mitochondrion contains (and indeed encodes) homologues of proteins specified by mtDNA. By contrast, although Dyall *et al.*⁷ and Hrady *et al.*⁸ also identified and studied two complex I homologues in hydrogenosomes, the genes encoding these two proteins (24 and 51 kDa) have not been found in any mtDNA to date but reside exclusively in the

nuclear genome. Admittedly, there is strong evidence that the mitochondrial 24- and 51-kDa subunits of complex I originate from the proto-mitochondrial genome via gene transfer to the nucleus. Nevertheless, their connection (and that of their hydrogenosomal counterparts, Ndh24 and Ndh51) to the proto-mitochondrion is less direct than in the case of proteins whose genes have been retained in at least some extant mitochondrial genomes.

Second, the mitochondrial affiliation demonstrated with Ndh51 by Hrady *et al.*⁸ is consistent with other data — particularly characteristics of the protein import system in hydrogenosomes — that unite these organelles with mitochondria^{2,3}. By contrast, there is no solid evidence that specifically affiliates the hydrogenosome of any anaerobic eukaryote with a different eukaryotic bacterial group, in particular an anaerobic hydrogen-producing lineage.

Finally, the sporadic phylogenetic distribution of hydrogenosomes and the intimate phylogenetic intermingling of their anaerobic ‘hosts’ with aerobic, mitochondrion-containing relatives imply that hydrogenosomes are derived secondarily from mitochondria. Indeed, it seems that nature can evolve a hydrogenosome from a mitochondrion with relative ease.

This story is far from complete, because mitochondria — putative remnant mitochondria that lack the ability to make ATP — have recently been discovered in several microbial lineages that do not have conventional mitochondria¹¹. The evolutionary and biochemical connections among mitochondria, hydrogenosomes and mitochondria must be elucidated if we are to truly understand the pathways and mechanisms of eukaryotic cell evolution. ■

Michael W. Gray is in the CIAR Program in Evolutionary Biology, and the Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 1X5, Canada. e-mail: m.w.gray@dal.ca

1. Müller, M. J. *Gen. Microbiol.* **139**, 2879–2889 (1993).
2. Embley, T. M. *et al. IUBMB Life* **55**, 387–395 (2003).
3. Dyall, S. D., Brown, M. T. & Johnson, P. J. *Science* **304**, 253–257 (2004).
4. Boxma, B. *et al. Nature* **434**, 74–79 (2005).
5. Gray, M. W., Burger, G. & Lang, B. F. *Science* **283**, 1476–1481 (1999).
6. Andersson, S. G. E., Karlberg, O., Canbäck, B. & Kurland, C. G. *Phil. Trans. R. Soc. Lond. B* **358**, 165–179 (2003).
7. Dyall, S. D. *et al. Nature* **431**, 1103–1107 (2004).
8. Hrdy, I. *et al. Nature* **432**, 618–622 (2004).
9. Clemens, D. L. & Johnson, P. J. *Mol. Biochem. Parasitol.* **106**, 307–313 (2000).
10. Akhmanova, A. *et al. Nature* **396**, 527–528 (1998).
11. Roger, A. J. & Silberman, J. D. *Nature* **418**, 827–829 (2002).

Atmospheric chemistry

The decay of organic aerosols

Euripides G. Stephanou

The chemistry of organic aerosols has been somewhat neglected on the assumption that they are eliminated from the atmosphere mainly by rainfall. Laboratory studies indicate that a rethink is called for.

Fine particles and droplets suspended in the atmosphere have a key role in environmental issues such as climate and human health. Over the oceans, such aerosols consist mainly of sulphates, but above continents they are mostly organic matter¹. Organic aerosols come from many sources, including smoke particles from burning fuels and biomass, and the light-induced oxidation of volatile hydrocarbons, both natural and man-made¹. The main process that removes organic aerosols from the atmosphere has been assumed to be precipitation, but writing in *Geophysical Research Letters* Molina and colleagues² suggest that another elimination route could be just as important.

Gaseous organic compounds in the atmosphere interact with oxidants such as ozone and hydroxyl and nitrate radicals, reactions that provide an important sink for their eradication from the atmosphere³. For organic aerosols, however, the most common means of removal is by deposition, either sedimentation — simply falling out of the atmosphere — or precipitation⁴. Organic aerosols are usually less than a micrometre in size⁵, so it is generally assumed⁴ that precipitation is the major process by which they leave the atmosphere; larger particles would be more likely to settle out. Molina and colleagues² now identify another removal pathway whereby the organic surface on atmospheric particles is degraded by oxidation initiated by hydroxyl radicals (OH[•]). The efficiency of this process appears to be comparable to precipitation in removing organic aerosols from the atmosphere.

To model the reactions that organic aerosols might undergo in the atmosphere, Molina *et al.*² used two organic films deposited on glass slides: a paraffin film to represent aliphatic aerosols (molecules with carbon chains); and a pyrene film to represent aromatic aerosols (having carbon-ring structures). Aliphatic and aromatic hydrocarbons such as paraffin and pyrene have

been isolated from organic aerosols from various locations⁶.

To examine the oxidation reactions of solid organic compounds, the authors exposed the model aerosol surfaces to an 'atmosphere' of various ratios of NO_x:O₂:H₂O, and then varied the concentration of OH[•] from 0.1 × 10⁸ to 100 × 10⁸ molecules per cm³ (the average global atmospheric OH[•] concentration⁷ is about 10⁶ molecules per cm³). Using state-of-the-art analytical instruments, they then measured the rate of degradation of the organic surface, how quickly the OH[•] is used up, and the type and speed of formation of the gaseous products.

Molina and colleagues clearly observed the loss of organic carbon from both model substrates. They also observed that, over time, the depletion rate of the organic layer is linearly dependent on the OH[•] concentration. The aromatic carbon surface degraded more slowly than the aliphatic one, suggesting that the route of decay varies according to the compound. The gaseous products of the degradation reaction are small, volatile, one- and two-carbon species; which particular species are produced depends on the substrate.

From their observations, the authors propose a mechanism for the OH[•]-induced oxidative degradation of organic aerosols. According to this, the reaction leads predominantly to a scission of the carbon-carbon bond in paraffin, and to cleavage of the aromatic ring in pyrene. The authors assume that the rate of carbon loss from the organic film is directly proportional to the OH[•] concentration, and, given an average OH[•] concentration of 10⁶ molecules per cm³, they estimate that an aliphatic aerosol of 0.02–0.2 μm will be converted entirely into gaseous products in about six days². The lifetime of an organic aerosol has been estimated from atmospheric measurements and lab experiments to be four to five days⁸. Consequently, this study concludes that oxidative degradation and removal by precipitation occur at comparable rates, and that OH[•]-induced oxidation is

a significant mechanism that eliminates organic aerosols from the atmosphere.

Chemical reactions with the OH[•] radical have been established as the dominant processes by which most gaseous organic compounds are removed from the atmosphere³. In fact, OH[•] reactions occur at environmentally significant rates even for the chemically recalcitrant PCBs (polychlorinated biphenyls), so they are an important atmospheric sink for these pollutants^{9,10}. However, there is only very limited information on the reactions of gaseous OH[•] with organic liquids and solids^{2,11}, or indeed on any of the chemistry of organic aerosols.

Traditional analytical techniques used to characterize organic aerosols failed to analyse the water-soluble organic compounds (which account for 70–90% of the aerosol mass¹²) and were limited to identifying only the components that could be dissolved in organic solvents (6–20% of aerosol mass¹²). We know now that the water-soluble fraction of total fine particulate aerosol mass contains oxygenated and macromolecular polar organic substances with surface-active properties¹². But the atmospheric chemistry of these polar species is otherwise relatively unknown and difficult to study. In addition, the association of some species, such as the environmental pollutants polyaromatic hydrocarbons, with black carbon particles seems to show a potential inhibiting effect for their reaction with gaseous OH[•] (ref. 11).

Molina and colleagues² make a strong case that the heterogeneous reactions of organic aerosols with atmospheric oxidants are important for their fate. The results highlight the need for further studies to improve our understanding of the reactions and effects of organic aerosols in the environment. We need a thorough chemical characterization and quantification of the main components, details of their reactions in the presence of atmospheric oxidants, and improved knowledge of their surface properties and water uptake before and after heterogeneous reactions in the atmosphere. Finally, lab experiments are rarely definitive, of course: systematic field studies will also be required. ■

Euripides G. Stephanou is in the Environmental Chemical Processes Laboratory, School of Sciences and Engineering, University of Crete, 71 409 Heraklion, Greece.

e-mail: stephanou@chemistry.uoc.gr

1. Andreae, M. O. & Crutzen, P. J. *Science* **276**, 1052–1058 (1997).
2. Molina, M. J. *et al. Geophys. Res. Lett.* **31**, L22104 (2004).
3. Atkinson, R. *et al. J. Phys. Chem. Ref. Data* **28**, 191–393 (1999).
4. Seinfeld, J. H. & Pandis, S. N. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change* (Wiley, New York, 1998).
5. Kavouras, I. G. & Stephanou, E. G. *J. Geophys. Res.* **107**, 4069–4080 (2002).
6. Rogge, W. F. *et al. Atmos. Environ.* **27A**, 1309–1330 (1993).
7. Prinn, R. G. *et al. Science* **292**, 1882–1888 (2001).
8. Liouss, C. *et al. J. Geophys. Res.* **101**, 19411–19432 (1996).
9. Anderson, P. N. & Hites, R. A. *Environ. Sci. Technol.* **30**, 1756–1763 (1996).
10. Mandalakis, M. *et al. Environ. Sci. Technol.* **37**, 542–547 (2003).
11. Esteve, W. *et al. Atmos. Environ.* **38**, 6063–6072 (2004).
12. Alves, C. *et al. J. Geophys. Res.* **107**, (D21) 8345 (2002).

An index of intactness

Georgina M. Mace

The global community is committed to reducing the rate of loss of biodiversity, but how can progress be measured? A novel system to tackle the problem may also identify key factors behind the changes.

Setting targets has become an increasingly common part of working life, and one that sometimes seems an unnecessary extra burden. But setting the target is just the beginning: gauging progress can be a major undertaking, and all this work will be in vain if the means to achieve the targets are not in place. In the case of biodiversity, measuring the ways in which different ecosystems are changing has proved a challenge¹, but on page 45 of this issue, Scholes and Biggs² unveil an innovative and practical approach that may also turn out to promote good management.

In 2002, the 188 countries that are signatories to the Convention on Biological Diversity committed themselves to “achieve by 2010 a significant reduction of the current rate of biodiversity loss at the global, regional and national level”³. Unfortunately, this laudable target is very vague as regards practicalities. It presents both a challenge and an opportunity for biodiversity scientists⁴: a challenge because biodiversity is not a simple concept, and coming up with measures that encompass all its aspects will be difficult; an opportunity because when such measures are in place, it will be possible to guide and manage biodiversity better, and so make progress towards a more sustainable world.

Scientists use the term ‘biodiversity’ to reflect almost every aspect of the living world, applying it across a range of spatial and temporal scales to encompass variability within and between genes, genomes, individuals, communities, traits and ecosystems, and including all organisms. Most policy-makers, in contrast, are used to seeing it represented simply as the changing number of species on a species list.

Evaluations of which aspects of biodiversity contribute to the health of an ecosystem clearly indicate that considering variability alone is not enough^{5–7}. Biodiversity assessments need to move away from a reliance on species lists and species extinction rates, because often the existence and proximity of local populations matters more.

Variability — the number or diversity of species in an area, say, or the number of genetic varieties of a crop strain in production — is necessary, but it is not sufficient to support the components of biodiversity that underlie key functions and benefits of an ecosystem. It is not hard to list circumstances where the quantity of a single component is crucial (for example, the biomass of forest

for timber, or the area of mangrove offering coastal protection), or where a species’ distribution in space and time is critical (for instance, pollinators need to be near their host species, and plant cover must be on valley sides to prevent erosion) (Fig. 1).

A systematic assessment of the dimensions of biodiversity — the different types (the number of different species, say), quantities and distributions at various ecological levels — will give a set of measurements. But it soon becomes clear that they are not all equal. Depending on the context and perspective, some are more significant than others (Fig. 1), and any meaningful evaluation of biodiversity will have to take account of this.

The development of appropriate global indicators for the 2010 target is progressing on a number of fronts. Existing data sets have been exploited to provide measures of forest area, protected area coverage, and trends in the abundance of certain species^{4,8,9}. Innovations and new data sets are revealing trends in the status of threatened species^{10,11}, and the geographical extent of additional ecosystems¹². But data to assess the full range of measures (Fig. 1) are extremely sketchy and unrepresentative because of the large gaps in our knowledge and the fact that there is little systematic monitoring. Genetic

measures across spatial scales are almost entirely missing. We have named and described fewer than 2 million of the 5 million to 30 million species expected to exist on Earth. Long-term monitoring covers only a tiny proportion of these, and is certainly unrepresentative. Even in relatively well-studied areas of the world, the number of biodiversity measures for which long-term trends can be assessed is remarkably limited. Clearly, new approaches are required if we are to make progress.

Scholes and Biggs’ biodiversity intactness index (BII)² makes a start in satisfying the many requirements, and provides a robust, sensitive but meaningful indicator. The index is built up from relative abundances of populations of species belonging to different taxonomic groups in different ecosystems, and facing different land-use management practices. It can be calculated for any political or geographical unit, and will give an indication of the overall condition of a region relative to a ‘pristine’ state. This state is defined by Scholes and Biggs as the unaltered, pre-industrial state, for which they use the current condition in protected areas as a surrogate measure.

Several features set their method apart from other available methods. The BII allows trends over time and space to be monitored readily. Also, and most usefully, because of the way it is constructed, the index can be separated out to provide comparative information across taxonomic groups, ecosystems or land-use management practices. Hence, unlike other methods that contribute to one measure of biodiversity (that is, one cell in Fig. 1), the BII can contribute to several at once. It can also assist in diagnosing the

Level	Importance of variability	Importance of quantity	Importance of distribution
Genes	Ultimate source of variability for evolution and adaptive change.	Influences evolution, affecting how new variants establish and spread through populations.	Different environments allow the evolution of local adaptation, resistance and resilience.
Species	Irreplaceable, unique units with combinations of traits from long and independent evolution. Intrinsic value.	Provisioning and regulating services may depend on quantity; e.g. food, fresh water. Long-term viability.	Local provisioning and regulating services; e.g. structural roles, pollinators. Community and ecosystem stability arises through the co-occurrence of species.
Populations	Local populations retain local adaptations.		
Ecosystems	Different ecosystems fulfil different roles.	Functions, products and services that depend on scale; e.g. protection from erosion, or volume of fresh water.	Functions, products and services that depend on location; e.g. fresh water near to communities that depend on it.

Figure 1 Measures of biodiversity. Across a range of levels at which biodiversity can be assessed, variability is not sufficient to capture the essential features that underpin the functioning and benefits of an ecosystem. Measures of both quantity and distribution are important too. The biodiversity intactness index devised by Scholes and Biggs² attempts to take such measures into account.

causes underlying an observed decline: changes can be traced back to reveal which taxonomic groups or ecosystems are losing populations of species the fastest, and whether the overall deterioration is due to many declining populations, a few localized extinctions, or a combination of the two.

The problem of data availability has been sidestepped rather than solved: Scholes and Biggs' calculation is based on expert opinion about how various species fare under different land use in each ecosystem. Clearly, real data would be preferable. But this method might also help to encourage the collection of data, because sampling systems established against this framework would be both achievable and useful, and might therefore be more likely to be implemented.

In addition, because land-use change is incorporated into the index, the results suggest where best to direct efforts to mitigate loss of biodiversity. For example, Scholes and Biggs' BILs for different taxa (Fig. 1 on page 47) show the relative sensitivity of birds, mammals and amphibians to a change in land use from moderate to degraded — that

is, use at a rate exceeding replenishment and causing widespread disturbance. Thus, this method has already moved beyond the stage of designing measures to suggesting actions to achieve the target. ■

Georgina M. Mace is at the Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK.

e-mail: georgina.mace@ioz.ac.uk

1. The Royal Society *Measuring Biodiversity for Conservation* (The Royal Society, London, 2003).
2. Scholes, R. J. & Biggs, R. *Nature* **434**, 45–49 (2005).
3. Convention on Biological Diversity www.biodiv.org/2010-target (2002).
4. Balmford, A. *et al. Science* **307**, 212–213 (2005).
5. Mace, G. M. *et al. in Ecosystems and Human Well-being: A Framework for Assessment* Vol. 1 (Millennium Ecosystem Assessment Ser., Island Press, Washington DC, 2005).
6. Luck, W. G., Daily, G. C. & Ehrlich, P. R. *Trends Ecol. Evol.* **18**, 331–336 (2003).
7. Balmford, A., Green, R. E. & Jenkins, M. *Trends Ecol. Evol.* **18**, 326–330 (2003).
8. Balmford, A., Crane, P., Dobson, A., Green, R. E. & Mace, G. M. *Phil. Trans. R. Soc. Lond. B* (in the press).
9. Loh, J. & Wackernagel, M. *Living Planet* (WWF Int., Gland, Switzerland, 2004).
10. Butchart, S. H. M. *et al. PLoS Biol.* **2**, e383 (2004).
11. Brooks, T. & Kennedy, E. *Nature* **431**, 1046–1047 (2004).
12. Convention on Biological Diversity www.biodiv.org/2010-target/indicators.aspx (2004).

Sonoluminescence

Cavitation hots up

Detlef Lohse

Gas inside collapsing bubbles can become very hot and, as a result, emit light. It turns out that temperatures of more than 15,000 kelvin can be reached — as hot as the surface of a bright star.

In 1917, Britain's Royal Navy had problems with bubble cavitation. This is a process in which tiny bubbles grow in size and then collapse as a result of pressure variations in the turbulent water around ships' propellers. The process is so violent that it was causing considerable damage to the propellers¹, so the navy asked the renowned physicist Lord Rayleigh to analyse the problem². His research led to what is now called the Rayleigh equation, which describes the dynamics of the collapsing bubble walls^{1,2}. However, the solution to the equation produced a singularity. It implied that, during collapse, the gas inside the bubble is compressed so fast that it cannot equilibrate with the surrounding liquid, leading to energy focusing and an infinite temperature increase. In reality, of course, this cannot happen, so the question is: what limits the temperature increase, or, in other words, how hot does the bubble get? On page 52 of this issue³, Flannigan and Suslick report a study of light emission from single bubbles during cavitation, and provide a direct answer to this question.

The temperature reached by the collapsing bubble depends on how much of the focused energy is lost by sound emission at the collapse

and how much is consumed by internal processes such as vibrations, rotations, dissociation and eventually ionization. If there are many collapsing bubbles, they disturb each other, which leads to a less-spherical collapse and therefore less-efficient energy focusing. Nonetheless, temperatures can rise so high that the bubbles start to glow. This phenomenon has already been investigated intensively by using sound waves to drive bubble production in liquids and then detecting the light emitted; the sound waves cause a temporarily reduced pressure in the liquid, which makes the bubbles grow and eventually collapse again (Fig. 1, overleaf). So far, emission spectra with a detailed line structure have only been observed for many transient bubbles together (so-called multi-bubble sonoluminescence). Analysis of the emitted spectral lines⁴ indicates that the temperature reached inside these bubbles is around 5,000 kelvin.

In single-bubble sonoluminescence^{5,6}, an isolated and stable bubble is studied; disturbances from other bubbles are absent. The light emission from such a bubble can be more than 10⁷ photons per flash⁷. As the bubble is driven periodically with sound waves at frequencies of typically 20–40 kHz, the emitted light is visible to the naked eye.



100 YEARS AGO

"Charge carried by the α Rays from Radium."
I have recently attacked this problem again, using the methods and apparatus previously described, but, in addition, employing a strong magnetic field to remove the slow-moving electrons present with the α particles. The apparatus was placed between the pole-pieces of an electromagnet, so that the field was parallel to the plane of the plates. In such a case, most of the escaping electrons describe curved paths and return to the plate from which they set out. On application of the magnetic field, a very striking alteration was observed in the magnitude of the current. The positive and negative currents for a given voltage were greatly reduced. The upper plate, into which the α particles were fired, rapidly gained a positive charge... I think these experiments undoubtedly show that the α particles do carry a positive charge, and that the previous failures to detect this charge were due to the masking action of the large number of slow-moving electrons emitted from the plates... Since the film of radium bromide is so thin that all the α particles escape from its surface, it is easy to deduce from the observed charge from a known weight of radium the total number of α particles expelled per second from one gram of radium bromide... a most important constant, for on it depends all calculations to determine the volume of the emanation, and of helium, the heat emission of radium, and also the probable life of radium and the other radio-elements.
E. Rutherford
From *Nature* 2 March 1905.

50 YEARS AGO

While recognizing the greatness of its opportunities and responsibilities in Europe, the [British] Council remarks: "It would be an exaggeration but not an untruth to say that a much closer understanding of the Englishman and his ways exists at Karachi than at Lyons, partly because Englishmen are a more familiar sight in one city than in the other, and partly because an outward similarity of culture helps to mask a basic difference of mental approach."... The Council exists as a body which helps to interpret overseas the permanent features of the British way of national life and to make available to the rest of the world the British contribution to knowledge, welfare or enjoyment.
From *Nature* 5 March 1955.

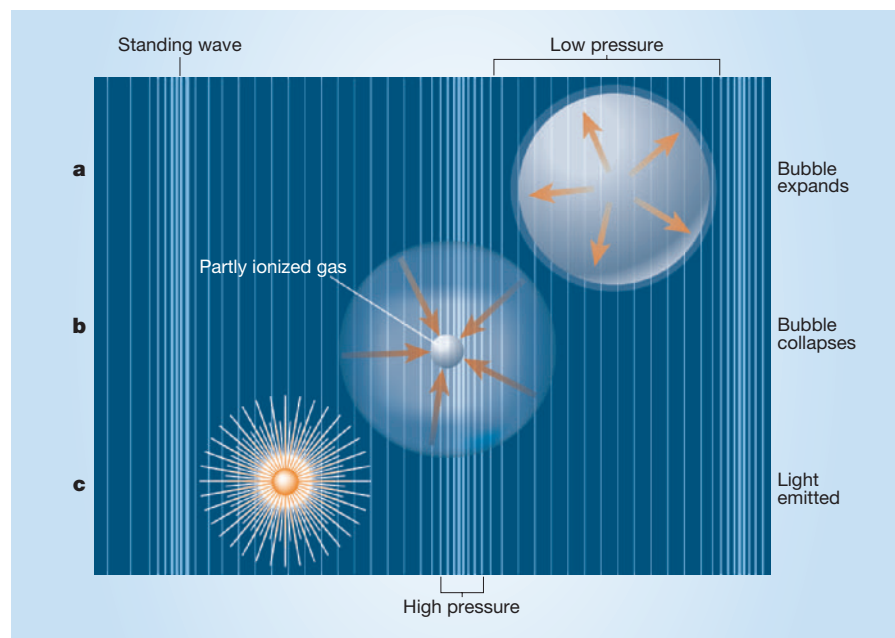


Figure 1 Bubble sonoluminescence — bubbles are driven by sound waves to emit light. **a**, At low sound-wave pressure, a gas bubble expands until **(b)** an increase in pressure triggers its collapse. Flannigan and Suslick³ find that, during collapse, temperatures can soar to 15,000 K, as the authors observed from spectra of light emitted from the bubble **(c)**. Analysis of the emission spectra also provides direct evidence for the existence of a plasma inside the collapsing bubbles.

However, it has previously been difficult to deduce the temperature reached, as the emission spectra from single bubbles were basically featureless.

But Flannigan and Suslick³ have obtained well-resolved spectral lines for the single-bubble case. They use xenon- and argon-filled bubbles in sulphuric acid, a set-up that has various advantages⁶. First, the high fluid viscosity of sulphuric acid ensures a stable spherical shape for relatively large bubbles. Second, monoatomic gases such as argon and xenon do not consume energy in rotational and vibrational degrees of freedom, and so more of the focused energy ends up as thermal energy. Third, because of the low vapour pressure of sulphuric acid, hardly any (polyatomic) vapour molecules invade the bubble at expansion; that would also eventually lead to additional energy absorption. In this way, Flannigan and Suslick are able to observe a thousand times more photons than observed from xenon and argon bubbles in water. As a result, they obtain good spectral details, from which a temperature of 15,000 kelvin is deduced — as high as is found at the surface of bright stars.

Perhaps an even more remarkable finding is that the emission spectra indicate the existence of plasma (ionized matter) inside the collapsing bubbles. Flannigan and Suslick observe that there are highly excited emissive states, which is inconsistent with thermal processes. Instead, some of the emitted light must originate from high-energy electrons and ions that are decelerated owing to collisions inside the gas bubble.

The presence of a weakly ionized plasma

and the origin of the light emission, as well as the high temperatures in single bubbles, have been predicted theoretically^{6,8–10}, but experimental evidence has been indirect. In previous work, the deduction of the bubble temperature from observable parameters required modelling assumptions (Fig. 2). Flannigan and Suslick's experiments are a milestone in single-bubble sonoluminescence, as they constitute the first direct measurement of the temperature

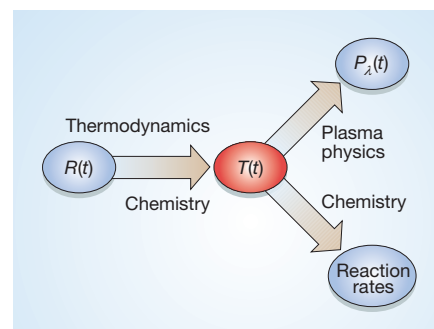


Figure 2 Indirect evidence for the temperature reached inside a collapsing bubble. Hitherto, the temperature $T(t)$ (as a function of time t) in single collapsing bubbles could only be deduced indirectly, using modelling steps to link observable parameters (blue circles) such as the chemical reaction rates¹¹, bubble radius $R(t)$, and the spectral radiance $P_A(t)$. Flannigan and Suslick³ have measured the temperature directly from light-emission spectral lines.

and the state of matter in a single bubble at collapse.

Detlef Lohse is in the Department of Applied Physics and the J. M. Burgers Center, University of Twente, 7500 AE, Enschede, The Netherlands.

e-mail: d.lohse@utwente.nl

1. Brennen, C. E. *Cavitation and Bubble Dynamics* (Oxford Univ. Press, 1995).
2. Rayleigh, L. *Phil. Mag.* **34**, 94–98 (1917).
3. Flannigan, D. J. & Suslick, K. S. *Nature* **434**, 52–55 (2005).
4. Flint, E. B. & Suslick, K. S. *Science* **253**, 1397–1399 (1991).
5. Crum, L. A. *Phys. Today* **47**, 22–29 (1994).
6. Brenner, M. P., Hilgenfeldt, S. & Lohse, D. *Rev. Mod. Phys.* **74**, 425–484 (2002).
7. Barber, B. P. *et al. Phys. Rep.* **281**, 65–144 (1997).
8. Moss, W. C. *et al. Science* **276**, 1398–1401 (1997).
9. Hilgenfeldt, S., Grossmann, S. & Lohse, D. *Nature* **398**, 402–405 (1999).
10. Toegel, R. & Lohse, D. *J. Chem. Phys.* **118**, 1863–1875 (2003).
11. Didenko, Y. T. & Suslick, K. S. *Nature* **418**, 394–397 (2002).

Cell cycle

Cyclin guides the way

Curt Wittenberg

The main enzymes that drive cell division can work on numerous substrates, but how is their specificity ensured? Regulatory subunits show the way, using various tricks to guide enzymes to their targets.

Even before Walther Flemming coined the term 'mitosis' in the 1880s, the choreography of cell division fascinated scientists. Since then it has become clear that the events that define different phases of the cell-division cycle are driven by distinct forms of an enzyme known as cyclin-dependent protein kinase.

Protein kinases facilitate the transfer of phosphate to protein substrates, generally altering their function or fate. As their name suggests, the cyclin-dependent kinases (CDKs) depend for their activity upon the binding of a regulatory subunit called a cyclin to the catalytic subunit. Many organisms use

numerous cyclins (and in some cases numerous CDKs) to drive the cell cycle. Different cyclin-CDK complexes phosphorylate different substrates and so have different effects. But how do cyclins influence the capacity of their catalytic partners to recognize substrates? On page 104 of this issue, Loog and Morgan¹ report that they can do so by altering the affinity of CDKs for their targets.

Structural complementarity between substrates and the active sites of enzymes — first proposed more than a hundred years ago by Emil Fischer in his 'lock-and-key' model — is in theory sufficient to account for the ability of the enzymes to discriminate

between potential substrates (Fig. 1a, b). But enzymes that modify proteins and other macromolecules need to distinguish between similar (or even identical) sites within larger, dissimilar molecules. To do so, they must recognize the differences between substrates. That problem has been solved by diversifying the task of target recognition (Fig. 1c). Whereas the motif to be modified (one or a few amino acids in a protein, for instance) is recognized by the enzyme's active site, discrimination between different substrates bearing that motif is often accomplished through specific interactions between other sites on the enzyme and substrate.

Cyclin-dependent kinases have apparently broken down this process even further. Whereas responsibility for recognizing the target motif (a serine or threonine followed by a proline) is delegated to a catalytic subunit (the CDK), both genetic and biochemical studies suggest that exchangeable regulatory subunits (the cyclins) have a role in discriminating between distinct protein substrates (Fig. 1d). This is, perhaps, best illustrated by baker's yeast (*Saccharomyces cerevisiae*), where the cell-cycle-regulatory CDK, called Cdk1, can associate with nine distinct cyclins — three G1 cyclins (Cln1–3) and six B-type cyclins (Clb1–6). These cyclins, in addition to activating Cdk1, direct it towards distinct biological outcomes.

But although cyclins had been implicated in substrate recognition, Loog and Morgan's paper¹ describes the first comprehensive study to compare the substrate specificity of purified CDK complexes that differ only in their cyclin. Their findings show that Clb5–Cdk1 and Clb2–Cdk1 complexes phosphorylate most members of a group of 150 previously confirmed Cdk1 substrates² with roughly equal efficiency. However, 26 of those substrates are phosphorylated 2.5–800 times as efficiently by Clb5–Cdk1. In contrast, Clb2–Cdk1 does not preferentially phosphorylate any of the proteins.

The authors go on to extend previous studies^{3–7} showing that a structural motif on the surface of some cyclins, referred to as the hydrophobic patch (HP), specifically interacts with a so-called RXL or Cy motif found on some CDK substrates and inhibitors. The HP motif is important for the biological activity of Clb5 (ref. 7). Loog and Morgan¹ now establish that this motif is essential for enhancing the activity of Clb5–Cdk1 towards its preferred substrates. Moreover, inactivating the Cy motif in the preferred Clb5–Cdk1 substrates eliminates their preferred status.

Strikingly, similar mutations in the Clb2 HP motif do not affect the efficiency with which Clb2–Cdk1 phosphorylates any of the substrates, regardless of the presence or absence of a Cy motif. That observation suggests that Clb2 does not use the HP motif for substrate recognition. In fact, Clb2 may not

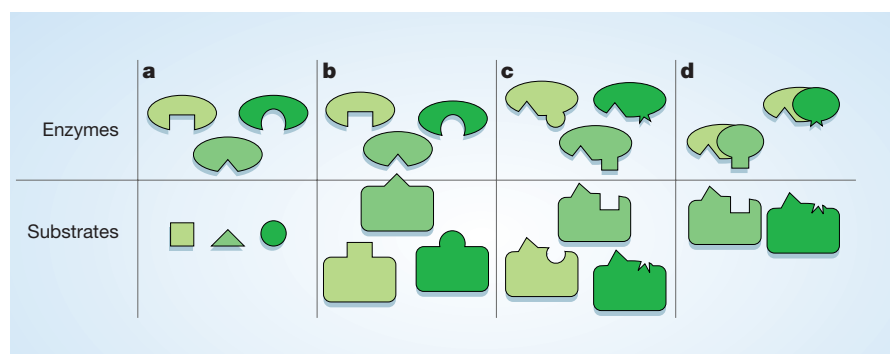


Figure 1 How enzymes select their substrates. **a, b**, In general, enzymes recognize their targets through structural complementarity between the substrate and the enzyme's active site (indicated here by the shape of the 'pocket'). Small substrates (**a**) and relatively small modification sites on proteins (**b**) can be recognized by this mechanism. **c**, Some enzymes make additional, specific contacts with the substrate that enable them to distinguish between proteins that have identical or related sites of modification. **d**, Loog and Morgan¹ have compelling new evidence that cyclin-dependent protein kinases (CDKs) have relegated that function to the exchangeable cyclin subunit, enabling a single CDK catalytic subunit to exist in numerous forms with different specificities.

confer substrate specificity upon Cdk1. It may simply activate it and leave substrate recognition entirely to the active site. In keeping with that interpretation, Archambault *et al.*⁸ have found that Cy-containing substrates depend upon the HP motif to interact with Clb5 in an *in vivo* assay, but that those lacking Cy motifs interact equally well with HP-deficient Clb5 and Clb2.

So what is the role of the HP motif in Clb2? Analysis of the relationship between the six yeast B-type cyclins reveals that, although Clb5 and Clb2 are closely related in terms of their overall sequence, their HP motifs appear to be significantly different⁸. Given the known structure of a complex between human cyclin A3 and a Cy-motif peptide³, the Clb2 HP motif seems to be incompatible with binding to the Cy motif⁸. Nevertheless, it has been well conserved between different organisms, suggesting that it is still important to Clb2's function. One possibility is that it regulates a function of Clb2–Cdk1 other than its enzymatic activity. Indeed, mutation of the HP motif in Clb2 impairs the protein's export from the nucleus and its localization to at least one site in the cytoplasm⁹. Because Loog and Morgan's analysis was performed largely *in vitro*, using purified proteins, the importance of subcellular localization in substrate selection was not evaluated.

Loog and Morgan's study¹ underlines the importance of cyclins in recognizing appropriate CDK substrates. The extent to which similar mechanisms are exploited by other cyclins remains to be fully examined, but there is ample evidence that other properties of cyclins are also important in substrate selection. Subcellular localization, already mentioned in the context of Clb2, is a well-established determinant of the biological function of yeast G1 cyclins^{10,11}. Of equal or even greater importance is the hallmark of

the cyclin proteins — their periodic accumulation during the cell cycle. Clearly, for a substrate to be phosphorylated it must be present in the cell along with the specific form of CDK that phosphorylates it.

So cyclins have a substantial role in directing CDKs to specific substrates. But there are numerous mechanisms for doing so, more than one of which may be used by a single cyclin. Ultimately, it is the combined action of these mechanisms that orchestrates the orderly progression of events leading to the faithful duplication of cells.

Curt Wittenberg is in the Departments of Molecular Biology and Cell Biology, Scripps Research Institute, La Jolla, California 92037, USA.

e-mail: curtww@scripps.edu

1. Loog, M. & Morgan, D. O. *Nature* **434**, 104–108 (2005).
2. Ubersax, J. A. *et al.* *Nature* **425**, 859–864 (2003).
3. Brown, N. R., Noble, M. E., Endicott, J. A. & Johnson, L. N. *Nature Cell Biol.* **1**, 438–443 (1999).
4. Kelly, B. L., Wolfe, K. G. & Roberts, J. M. *Proc. Natl Acad. Sci. USA* **95**, 2535–2540 (1998).
5. Sorensen, C. S. *et al.* *Mol. Cell Biol.* **21**, 3692–3703 (2001).
6. Takeda, D. Y., Wohlschlegel, J. A. & Dutta, A. *J. Biol. Chem.* **276**, 1993–1997 (2001).
7. Wilmes, G. M. *et al.* *Genes Dev.* **18**, 981–991 (2004).
8. Archambault, V., Buchler, N. E., Wilmes, G. M., Jacobson, M. D. & Cross, F. R. *Cell Cycle* **4**, 125–130 (2005).
9. Bailly, E., Cabantous, S., Sondaz, D., Bernadac, A. & Simon, M. N. *J. Cell Sci.* **116**, 4119–4130 (2003).
10. Edgington, N. P. & Futcher, B. *J. Cell Sci.* **114**, 4599–4611 (2001).
11. Miller, M. E. & Cross, F. R. *Mol. Cell Biol.* **20**, 542–555 (2000).

Correction

A misleading statement appeared in the News and Views article "Cardiology: Solace for the broken-hearted?" by Christine L. Mummery (*Nature* **433**, 585–587; 2005). The cardiac arrhythmias reported in reference 9 (P. Menasché *et al.*, *J. Am. Coll. Cardiol.* **41**, 1078–1083; 2003) were not the cause of fatalities in patients who received their own skeletal-muscle progenitor cells as therapy for heart damage, as implied in the passage concerned.

High-pressure physics

Hydrogen bonds in symmetry

Phys. Rev. Lett. **94**, 065505 (2005)

Squeezing crystalline formic acid (HCOOH) turns it into a polymer, Alexander F. Goncharov *et al.* report.

They find that at about 20 gigapascals (GPa) the hydrogen bonds linking formic acid molecules into chains become symmetric: the hydrogen atoms sit precisely midway between two oxygen atoms on neighbouring molecules. This means that the hydrogens no longer unambiguously 'belong' to one molecule or the other. It also puts the hydrogen atoms in a decidedly 'non-classical' environment, in which they form chemical bonds to two atoms rather than just one. These symmetric hydrogen bonds, the researchers' quantum-chemical calculations show, have a partially covalent character. Symmetric hydrogen bonds have been proposed to exist in the high-pressure phase of ice known as ice X, but experimental support for them has remained contentious.

Above 40 GPa, Goncharov *et al.* find that long-range order in their formic acid samples disappears: the solid becomes amorphous, as the hydrogen-bridged chains become fully fledged polymers. This polymeric form persists when the pressure is relaxed, until about 20 GPa.

Philip Ball

Microtechnology

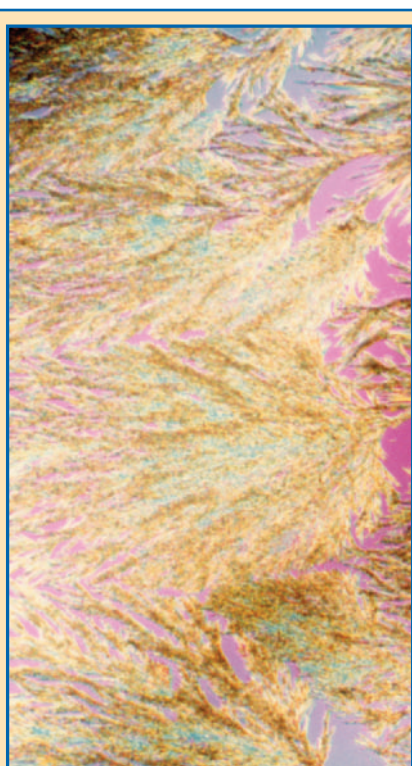
Turn on

Small **1**, 202–206 (2005)

It's an age-old question for engineers: how can a power source be translated into rotational motion? Jeffrey M. Catchmark and colleagues have been investigating how an existing system of nanorods made of gold at one end and platinum at the other, bathed in a solution of hydrogen peroxide, might be adapted to that purpose.

The platinum catalyses the production of oxygen, which results in a concentration gradient — and so an interfacial tension gradient — along the gold segment of the nanorod. As a result, the rod is propelled forward. An essential feature of the system is that the gold surface is hydrophobic.

To harness this power production for rotational motion, Catchmark *et al.* manufactured gear-like structures, 150 μm in diameter, made of gold and with platinum implants on one edge of the gear teeth. When fuelled by a hydrogen peroxide solution, the gears refused to spin. The trick, it turned out, was to add small amounts of acids to the solution, making the gold surface hydrophobic and producing motion at a rate of about one rotation per second —



Taste

Bitter variations

Curr. Biol. **15**, 322–327 (2005)

We truly do inherit our tastes from our parents. For example, the reason that only some people perceive the compound phenylthiocarbamide (PTC) as bitter has a genetic basis. Work by Bernd Bufe and colleagues now shows why those who find it bitter do so to vastly different degrees.

The gene encoding the PTC taste receptor was identified last year, but no functional variations of the bitter receptor had been identified that would account for the taste differences. Bufe *et al.* now link specific versions of the gene — alleles — to levels of PTC perception. They introduced different alleles into cultured cells, which expressed the taste-receptor protein for the compound. By measuring the response of the receptors to PTC (shown here in crystal form), Bufe *et al.* identified which alleles conveyed the greatest cellular sensitivity.

Further tests on human subjects enabled the authors to confirm which versions of the PTC bitter-taste gene give people the greatest ability to detect the compound. At the gastronomic level, the findings provide a molecular context for individual fussiness over foods such as broccoli.

Roxanne Khamsi

in linear terms, much faster than previously observed nanorod velocities.

From other experiments, the authors confirm that the acids maintain hydrophobicity at the gold surface. But there is evidently much still to learn about the surface chemistry involved, and the gears have yet to be coupled up mechanically.

Tim Lincoln

Animal behaviour

Snap responses

Biol. Lett. doi:10.1098/rsbl.2004.0237 (2005)

Eusocial animals exhibit such characteristics as division of reproductive labour between castes, cohabiting generations and cooperative behaviour. There are plenty of examples among the insects; naked mole rats are an instance among the vertebrates.

Some species of shrimp also show eusocial characteristics. Eva Tóth and J. Emmett Duffy now describe a further aspect of shrimp eusociality, that of a collective response of members of a colony in the face of threats. They looked at species of *Synalpheus*, tiny inhabitants of sponges in the tropics that are fiercely territorial and mark their displeasure by snapping their 'fighting claw'.

Tóth and Duffy observed the behaviour of *Synalpheus* when confronted by an alien shrimp of the same species. The initial one-to-one confrontation elicited a snap response from the defender. But if the intruder was brazen enough to push its luck, other colony members joined in with a cacophony of snapping. The aim of this collective sabre-rattling, say the authors, is not to enlist physical help against attack but to provide an unequivocal signal that the sponge is already colonized.

Tim Lincoln

Chemical biology

Pore sequencing

Angew. Chem. Int. Edn **44**, 1401–1404 (2005)

Reading the sequence of a single strand of DNA by pulling it through a tiny pore one base at a time may be feasible, Nurit Ashkenasy *et al.* show. They find that the ion current through a natural pore-forming protein, α -haemolysin, embedded in a lipid membrane, depends on which kind of DNA base is lodged at a key position in the pore neck: a deoxyadenosine (dA) group in this position produces a different signal from a deoxycytosine (dC) group.

It has been proposed previously that this kind of base discrimination could be used for rapid, single-molecule gene sequencing, and single DNA strands containing just purine bases (such as poly-A) have been differentiated from pyrimidine-only (poly-C) strands in conductivity measurements. But can the technique identify bases one at a time?

Ashkenasy *et al.* have made DNA single strands 'knotted' at one end with a hairpin turn, so that they cannot pass right through α -haemolysin but get lodged in the pore. They find that a single A base in a poly-C sequence can be distinguished if it is precisely 20 nucleotides away from the base of the hairpin loop, but not if it is at positions 19 or 21. Thus, there is a critical 'reading site' within the protein channel that makes this form of sequencing possible in principle.

Philip Ball

Postprandial cardiac hypertrophy in pythons

This snake can synthesize fresh heart muscle to cope with extra metabolic demand.

Oxygen consumption by carnivorous reptiles increases enormously after they have eaten a large meal in order to meet metabolic demands, and this places an extra load on the cardiovascular system. Here we show that there is an extraordinarily rapid 40% increase in ventricular muscle mass in Burmese pythons (*Python molurus*) a mere 48 hours after feeding, which results from increased gene expression of muscle-contractile proteins. As this fully reversible hypertrophy occurs naturally, it could provide a useful model for investigating the mechanisms that lead to cardiac growth in other animals.

The heart is remarkable for its ability to remodel itself in response to altered functional demands. For example, chronic exercise training in mammals results in ventricular hypertrophy¹, which is beneficial because the resulting increase in stroke volume leads to a decrease in the resting and submaximal heart rates, and to an increase in filling time and in venous return².

Burmese pythons are considered to be an excellent model of extreme physiological upregulation³. While digesting, their metabolic rate may increase by up to 40-fold relative to the fasting rate and may be raised for as long as 14 days (ref. 3). This increase in oxygen consumption is accompanied by rapid remodelling of many physiological systems: within two days of feeding, there is a substantial increase in wet mass of the gastrointestinal system, kidneys, liver, pancreas, lungs, heart and stomach⁴. However, the cause of this remodelling, which could be increased protein synthesis or increased fluid content, is unclear^{3,5}.



Hearty meal: the python's cardiac ventricles grow after feeding.

To investigate the nature of cardiac hypertrophy in pythons following feeding, we obtained ventricles from three groups of Burmese pythons: fasting (fast of 28 days), digesting (two days after consuming rats equal to $25.0 \pm 0.1\%$ body mass) and post-digestion (28 days after the meal). (For methods, see supplementary information.) Oxygen consumption increased sevenfold and ventricular mass increased significantly ($P < 0.003$) by 40% during digestion (Table 1). This increase was fully reversible, as the ventricular mass returned to its fasting mass in post-digestion animals.

There was no change in the ventricular dry/wet mass ratio, indicating that the increased ventricular mass during digestion was not due to water shifts between extra- and intracellular compartments. Total protein, RNA and myofibrillar concentrations on a tissue-mass-specific basis did not change during digestion (Table 1).

Mass-specific DNA concentration significantly decreased ($P < 0.01$), and this is consistent with the ventricular mass increase by cellular hypertrophy found in rats⁶. All of these measurements indicate a rapid new growth of ventricular tissue.

To investigate whether this growth was a result of increased transcription, we sequenced the isoforms of cardiac myosin heavy chains (GenBank accession numbers: AY773093 and AY773094). We found a significant increase in the expression of messenger RNA for heavy-chain cardiac myosin during digestion (Table 1), judging both by polymerase chain reaction with reverse transcription ($P < 0.0001$) and by band intensity on northern blots

($P < 0.0001$; Fig. 1).

We conclude that the newly synthesized protein results from increased transcription of the gene encoding cardiac myosin heavy chains and that cardiac hypertrophy follows from *de novo* addition of contractile elements. This cardiac hypertrophy is likely to have important consequences for oxygen transport and could explain why stroke volume in postprandial pythons is 50% greater than that measured in fasted animals doing maximal exercise⁷.

This ventricular growth in postprandial pythons is very rapid compared with that in mammalian models, in which comparable increments in ventricular size take weeks to develop⁸. In addition, mammalian models may necessitate highly invasive procedures for variable aortic occlusion, such as hydraulic constrictors, inflatable cuffs or angioplasty balloons, which can induce acute congestive failure and aortic rupture

Table 1 Cardiac hypertrophy following digestion in Burmese pythons

Phase	Fasting	Digesting	Post-digestion
Pre-feeding body mass (kg)	0.702 $\pm$ 0.045	0.783 $\pm$ 0.126	0.721 $\pm$ 0.070
Oxygen consumption (ml O ₂ min ⁻¹ kg ⁻¹)	0.67 $\pm$ 0.056	4.49 $\pm$ 0.385 *	0.54 $\pm$ 0.070
Ventricular mass (g wet mass)	0.95 $\pm$ 0.070	1.34 $\pm$ 0.202 *	1.07 $\pm$ 0.062
Ventricular mass (% body mass)	0.136 $\pm$ 0.006	0.168 $\pm$ 0.004 *	0.137 $\pm$ 0.007
Dry/wet mass of ventricles (%)	22.4 $\pm$ 1.470	23.1 $\pm$ 1.91	20.3 $\pm$ 0.06
Total protein (μ g per mg ventricle)	236 $\pm$ 17	208 $\pm$ 10	249 $\pm$ 24
DNA (ng per mg ventricle)	2.93 $\pm$ 0.11	2.54 $\pm$ 0.12 *	3.09 $\pm$ 0.12
RNA (μ g per mg ventricle)	0.96 $\pm$ 0.064	1.11 $\pm$ 0.080	1.08 $\pm$ 0.080
Myofibrillar protein (μ g per mg ventricle)	32.9 $\pm$ 2.25	34.8 $\pm$ 3.78	30.7 $\pm$ 2.89
Ventricular Mhc mRNA (Mhc/18S)	0.97 $\pm$ 0.163	2.11 $\pm$ 0.221 *	0.78 $\pm$ 0.139
Ventricular Mhc mRNA (intensity)	2,320 $\pm$ 526	31,650 $\pm$ 5,280 *	5,140 $\pm$ 1,020 *

Mhc, cardiac myosin heavy chain; fasting, 28 d fasted; digesting, 48 h postprandial; post-digestion, 28 d postprandial. Ventricular Mhc mRNA expression was determined by semiquantitative reverse-transcription polymerase chain reaction (RT-PCR) and by northern blot (Fig. 1). In RT-PCR, Mhc mRNA is shown as a ratio (Mhc/18S) of 18S ribosomal subunit mRNA expression, which is considered to be constant. Quantitative mRNA expression measured by northern blot is given as intensity. $N = 6$ per group. Values are means $\pm$ 1 s.e.m. All experiments were carried out under University of California at Irvine Animal Research Committee Protocol Number 1999-2123.

*Significantly different ($P < 0.05$) from fasting value by one-tail t -test.

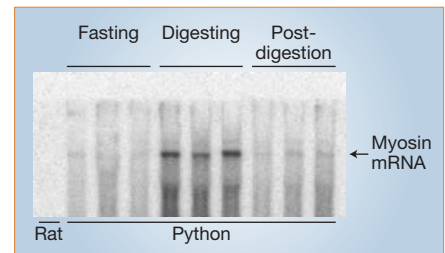


Figure 1 Northern blot showing increase in myosin mRNA in the cardiac ventricles of postprandial Burmese pythons. Each lane represents RNA from an individual python ventricle, hybridized to a probe specific to the ventricular myosin heavy-chain isoform. The same blot hybridized with a probe for the python atrial isoform gave no detectable signal. The lane immediately to the left of the fasted samples contains rat ventricular RNA, showing that the probe does not crossreact. Fasting phase, 28 d fasted; digesting phase, 48 h postprandial; post-digestion phase, 28 d postprandial.

as well as hypertrophy. Because Burmese pythons naturally undergo a 40%, fully reversible increase in ventricular mass in the two days after a meal, they could provide an attractive model for investigating the fundamental mechanisms that lead to cardiac remodelling and ventricular growth⁹. The physiological stimuli underlying this hypertrophy are still unknown, but are likely to include neural and humoral factors.

Johnnie B. Andersen*, Bryan C. Rourke†, Vincent J. Caiozzo‡, Albert F. Bennett*, James W. Hicks*

Departments of *Ecology and Evolutionary Biology, and ‡Orthopaedics, University of California, Irvine, California 92697, USA

e-mail: jhicks@uci.edu

†Department of Biological Sciences, California State University, Long Beach, California 90840, USA

- Cooper, G. *Annu. Rev. Physiol.* **49**, 501–518 (1987).
- Richey, P. A. & Brown, S. P. *J. Sports Sci.* **16**, 129–141 (1998).
- Secor, S. M. & Diamond, J. *Nature* **395**, 659–662 (1998).
- Secor, S. M. & Diamond, J. *Am. J. Physiol.* **272**, R902–R912 (1997).
- Starck, J. M. & Beese, K. *J. Exp. Biol.* **204**, 325–335 (2001).
- Vliegen, H. W., Bruschke, A. V. G. & Van Der Laarse, A. *Comp. Biochem. Physiol. A: Physiol.* **95**, 109–114 (1990).
- Secor, S. M., Hicks, J. W. & Bennett, A. F. *J. Exp. Biol.* **203**, 2447–2454 (2000).
- Morgan, H. E. *et al. Annu. Rev. Physiol.* **49**, 533–543 (1987).
- Quinn, K. E. *et al. Biophys. J.* **74**, A355 (1998).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Synaesthesia

When coloured sounds taste sweet

Synaesthesia is the involuntary physical experience of a cross-modal linkage — for example, hearing a tone (the inducing stimulus) evokes an additional sensation of seeing a colour (concurrent perception). Of the different types of synaesthesia, most

have colour as the concurrent perception¹, with concurrent perceptions of smell or taste being rare^{2,3}. Here we describe the case of a musician who experiences different tastes in response to hearing different musical tone intervals, and who makes use of her synaesthetic sensations in the complex task of tone-interval identification. To our knowledge, this combination of inducing stimulus and concurrent perception has not been described before.

E.S. is a 27-year-old professional musician who is female, right-handed and of average intelligence⁴ (IQ, 115). Whenever she hears a specific musical interval, she automatically experiences a taste on her tongue that is consistently linked to that particular interval (Table 1). Besides this exceptional interval-to-taste synaesthesia, she also reports the more common tone-to-colour synaesthesia, in which each particular tone is linked to a specific colour (for example, C and red; F sharp and violet).

Both synaesthetic perceptions have always been consistently reproducible. We repeatedly tested E.S. for over a year and have confirmed that her interval-to-taste synaesthesia is unidirectional: she does not hear tone intervals when exposed to taste. In addition, E.S. applies this synaesthesia in identifying tone intervals (which is evidence of a synaesthesia–cognition cascade).

To assess the influence of E.S.'s synaesthetic gustatory perception on her ability to identify tone intervals, we adapted the Stroop task⁵ (for methods, see supplementary information). Four selected tone intervals (seconds and thirds) were presented while applying four differently tasting solutions (sour, bitter, salty and sweet) to E.S.'s tongue. Her task was to identify the tone intervals by pressing a particular button for each interval on a computer keyboard. Reaction times and errors were measured for trials in which the applied taste was either congruent or incongruent with the tone interval; tone intervals were also presented without taste stimulation.

We found that E.S.'s tone-interval identification was perfect and was significantly faster during the congruent condition compared with all the other conditions (Fig. 1). Five non-synaesthetic musicians were tested as controls using the same procedure: no significant between-condition differences were found. The reaction times of the controls were comparable to those of E.S. in the no-taste condition (Fig. 1).

To exclude conceptual priming effects as an explanation for these results (for example, the subject might imagine sourness when presented with 'sour' as either a taste or word), we also tested E.S. by showing her the word(s) describing each taste. We found no between-condition difference in this conceptual task (Fig. 1).

Together, these results indicate that E.S.'s

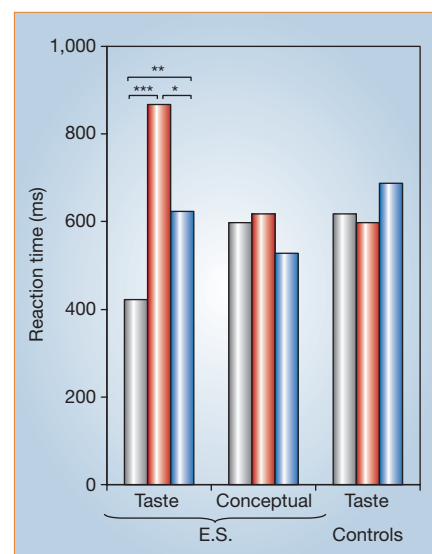


Figure 1 Mean reaction times in a gustatory Stroop task linking perception of tone intervals with different tastes for congruent-taste (grey), incongruent-taste (red) and no-taste (blue) conditions for synaesthete E.S. and for five non-synaesthetic musicians (controls). In the 'Taste' condition, musical intervals were presented while solutions of different taste (citric acid, 20 g litre⁻¹; quinine, 60 mg litre⁻¹; salt, 10 g litre⁻¹; sucrose, 120 g litre⁻¹) were delivered to the subject's tongue. The 'Conceptual' condition followed the same procedure, except that words describing the tastes, instead of the tastes themselves, were visually presented 2 s before the tone interval. Non-parametric randomization tests were used for statistical comparison. For E.S., all statistical comparisons in the taste condition were associated with *P* values of less than 0.01 (**P* < 0.05, ***P* < 0.01, ****P* < 0.001). For control subjects and for the conceptual condition, none of the comparisons revealed significant differences. The reaction time of E.S. in the no-taste condition is similar to those of the controls, but is faster in the congruent condition and slower in the incongruent condition.

performance in the gustatory Stroop task is most likely to be due to her extraordinary type of synaesthesia, in which a complex inducing stimulus leads to a systematic, concurrent gustatory sensation. This case differs from another gustatory synaesthete, S., who reported blended gustatory sensations (such as specific meals) in response to simple auditory stimuli (tones and sounds)². E.S.'s application of her synaesthetic sensations in identifying tone intervals — a complex task that requires formal musical training — demonstrates that synaesthetics may be used to solve cognitive problems.

Gian Beeli, Michaela Esslen, Lutz Jäncke

Institute of Neuropsychology, University of Zurich, 8032 Zurich, Switzerland

e-mail: l.jaencke@psychologie.unizh.ch

- Rich, A. N. & Mattingley, J. B. *Nature Rev. Neurosci.* **3**, 43–52 (2002).
- Luria, A. R. *The Mind of a Mnemonist* (Basic Books, New York, 1969).
- Ward, J. & Simner, J. *Cognition* **89**, 237–261 (2003).
- Horn, W. *Leistungsprüfungssystem* (Hogrefe, Göttingen and Bern, 1983).
- Stroop, J. R. *J. Exp. Psychol.* **18**, 643–662 (1935).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Table 1 Tastes triggered by tone intervals

Tone interval	Taste experienced
Minor second	Sour
Major second	Bitter
Minor third	Salty
Major third	Sweet
Fourth	(Mown grass)
Tritone	(Disgust)
Fifth	Pure water
Minor sixth	Cream
Major sixth	Low-fat cream
Minor seventh	Bitter
Major seventh	Sour
Octave	No taste

Tastes experienced by synaesthete E.S. in response to different musical tone intervals are shown; in the case of the fourth and tritone intervals, however, complex visual and emotional perceptions, respectively, are induced. Note that dissonant tone intervals induce unpleasant tastes and consonant ones induce pleasant ones (for example, the minor second intervals induce sour tastes, and the major thirds induce sweet ones). There is also an apparent symmetry in some of the responses: the minor second and major seventh, which are mirror-image intervals in terms of octave equivalence, are both rated as sour, and the major second and minor seventh are both rated as bitter.

Quantum computing with realistically noisy devices

E. Knill

Mathematical and Computational Sciences Division, National Institute of Standards and Technology, Boulder, Colorado 80305, USA

In theory, quantum computers offer a means of solving problems that would be intractable on conventional computers. Assuming that a quantum computer could be constructed, it would in practice be required to function with noisy devices called ‘gates’. These gates cause decoherence of the fragile quantum states that are central to the computer’s operation. The goal of so-called ‘fault-tolerant quantum computing’ is therefore to compute accurately even when the error probability per gate (EPG) is high. Here we report a simple architecture for fault-tolerant quantum computing, providing evidence that accurate quantum computing is possible for EPGs as high as three per cent. Such EPGs have been experimentally demonstrated, but to avoid excessive resource overheads required by the necessary architecture, lower EPGs are needed. Assuming the availability of quantum resources comparable to the digital resources available in today’s computers, we show that non-trivial quantum computations at EPGs of as high as one per cent could be implemented.

Research in quantum computing is motivated by the great increase in computational power offered by quantum computers^{1–3}. There is a large and still-growing number of experimental efforts whose ultimate goal is to demonstrate scalable quantum computing. Scalable quantum computing requires that arbitrarily large computations can be implemented efficiently with little error in the output. Criteria that need to be satisfied by devices used for scalable quantum computing have been specified⁴. An important one is that the level of noise affecting the physical gates and memory is sufficiently low. The type of noise affecting a given implementation is the ‘error model’. A scheme for scalable quantum computing in the presence of noise is a ‘fault-tolerant architecture’. In view of the low-noise criterion, studies of scalable quantum computing involve constructing fault-tolerant architectures and providing answers to questions such as the following: is scalable quantum computing possible for error model $\mathcal{E}$? Can fault-tolerant architecture $\mathcal{A}$ be used for scalable quantum computing with error model $\mathcal{E}$? What resources are required to implement quantum computation $\mathcal{C}$ using fault-tolerant architecture $\mathcal{A}$ with error model $\mathcal{E}$?

To obtain broadly applicable results, fault-tolerant architectures are constructed for generic error models. Here, the error model is parametrized by an error probability per gate (or simply error per gate, EPG), where the errors are unbiased and independent. The fundamental theorem of scalable quantum computing is the threshold theorem and answers the first question as follows: if the EPG is smaller than a threshold, then scalable quantum computing is possible^{5–8}. Thresholds depend on additional assumptions for the error model and device capabilities. Estimated thresholds vary from below 10^{-6} (refs 5–8) to 3×10^{-3} (ref. 9), with 10^{-4} (ref. 10) often quoted as the EPG to be achieved in experimental quantum computing.

In the few cases where experiments with two quantum bits (qubits) have been performed, the EPGs currently achieved are much higher, 3×10^{-2} or more in ion traps^{11,12} and liquid-state nuclear magnetic resonance^{13,14}. For quantum computing to become practical, it is essential to reduce the large gap between the experimentally achieved EPGs and those required by theory. The first goal of our work is to provide evidence that scalable quantum computing is possible at EPGs above 3×10^{-2} . This is encouraging, but the fault-tolerant architecture that achieves this is impractical because of its large resource requirements. To reduce the resource requirements, lower EPGs are required. The second goal of our work is to give a simple fault-tolerant architecture (called the ‘ C_4/C_6

architecture’) well-suited to efficient computing with EPGs between 10^{-4} and 10^{-2} . The third goal is to provide a means of estimating its resource requirements depending on computation size and EPG.

Fault-tolerant architectures realize low-error qubits and gates by encoding them with error-correcting codes. A standard technique for reducing errors is concatenation. Suppose we have a scheme that, starting with qubits and gates at one EPG, produces encoded qubits and gates that have a lower EPG. If the error model for encoded gates is sufficiently well-behaved, we can apply the same scheme to the encoded qubits and gates to obtain a next level of encoded qubits and gates with much lower EPGs. This process yields a hierarchy of repeatedly encoded qubits and gates, where the physical qubits and gates are at level 0. The top level is used for quantum computing. Its qubits, gates, EPGs and so on are ‘logical’.

The C_4/C_6 architecture differs from previous ones by combining a number of independently useful techniques. First, we use the simplest error-detecting codes, thus avoiding the complexity of even the smallest error-correcting codes. Error correction is added naturally by concatenation. Second, error correction is performed in one step and combined with logical gates by means of error-correcting teleportation. This minimizes the number of gates contributing to errors before they are corrected. Third, the architecture bootstraps key gates by state preparation and purification, thus enabling us to define it using a minimal and incomplete set of operations with only one unitary gate. Fourth, verification of the needed ancillary states (logical Bell pairs) largely avoids the traditional syndrome-based schemes. Instead, we use hierarchical teleportations and partial decoding. Finally, the highest thresholds are obtained by introducing the model of postselected computing with its own thresholds, which may be higher than those for standard quantum computing. Our fault-tolerant implementation of postselected computing has the property that it can be used to prepare states that are sufficient for standard scalable quantum computing.

Error model and assumptions

The unit of quantum information is the qubit, a quantum two-level system whose states are superpositions $\alpha|0\rangle + \beta|1\rangle$ (ref. 15). Qubits are acted on by the Pauli operators $X = \sigma_x$ (bit flip), $Z = \sigma_z$ (sign flip) and $Y = \sigma_y = i\sigma_x\sigma_z$. The identity operator is I . One-qubit gates include preparation of $|0\rangle$ and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$, Z -measurement (distinguishing between $|0\rangle$ and $|1\rangle$), X -measurement (distinguishing between $|+\rangle$ and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$), and

the Hadamard gate ($\text{HAD}, \alpha|0\rangle + \beta|1\rangle \mapsto \alpha|+\rangle + \beta|-\rangle$). We use one unitary two-qubit gate, the controlled-NOT (CNOT), which maps $|00\rangle \mapsto |00\rangle$, $|01\rangle \mapsto |01\rangle$, $|10\rangle \mapsto |11\rangle$ and $|11\rangle \mapsto |10\rangle$. This set of gates is a subset of the so-called Clifford gates, which are insufficient for universal quantum computing¹⁰. Our minimal gate set $\mathcal{G}_{\min}$ consists of $|0\rangle$ and $|+\rangle$ preparation, Z - and X -measurement and CNOT. Universality may be achieved with the addition of other one-qubit preparations or measurements, as explained below. The physical gates mentioned are treated as being implemented in one ‘step’; the actual implementations may be more complex.

All error models can be described by inserting errors acting as quantum operations (not necessarily unitary) after gates or before measurements. We assume that a gate’s error consists of random, independent applications of products of Pauli operators with probabilities determined by the gate. It is at present too difficult to obtain a threshold that does not depend on the details of the probability distributions, so we assume unbiased, ‘depolarizing’ errors for each gate: $|0\rangle$ ($|+\rangle$) state preparation erroneously produces $|1\rangle$ ($|-\rangle$) with probability e_p . A binary (such as Z or X) measurement results in the wrong outcome with probability e_m . CNOT is followed by one of the 15 possible non-identity Pauli products acting on its qubits with probability $e_c/15$ each. HAD is modified by one of the Pauli operators, each with probability $e_h/3$. We further simplify by setting $e_c = \gamma$, $e_h = 4\gamma/5$ and $e_m = e_p = 4\gamma/15$. This choice is justified as follows: $4\gamma/15$ is the one-qubit marginal probability of error for the CNOT, which we expect to be an upper bound for all one-qubit gate errors. As for preparation errors, if they are much larger than $4\gamma/15$, then it is possible to purify prepared states using a CNOT. For example, we can prepare $|0\rangle$ twice, apply a CNOT from the first to the second and measure Z of the second. We then try again if the measurement outcome indicates $|1\rangle$, and otherwise use the first state. The probability of error is given by $4\gamma/15 + O(\gamma^2)$, assuming that CNOT error is as above and measurement and preparation errors are proportional to γ . To improve Z measurement, we can introduce an ancilla in $|0\rangle$, apply a CNOT from the qubit to be measured to the ancilla, and measure both qubits. If the measurements disagree, an error is detected. If not, the conditional measurement error probability is $4\gamma/15 + O(\gamma^2)$. Detected errors are readily managed¹⁶.

To account for ‘memory’ errors, we assume that gates other than measurements take the same amount of time. Thus, the error parameters represent the total error, including any delays for faster gates to equalize gate times. For the C_4/C_6 architecture, memory is

required when waiting for measurement outcomes that determine whether prepared states are good, or that are needed after teleportation, particularly when implementing non-Clifford gates^{17,18}. The simplest way to account for memory errors in these situations is to distribute it equally to the surrounding gates. The maximum error thus distributed is the memory error e_B accumulated during the time required for a Bell measurement, here consisting of a CNOT followed by X and Z measurements. No gate is both preceded and followed by memory delays, so gate errors are adjusted by at most $e_B/2$, which we assume is already taken into account in the errors given in the previous paragraph. To ensure that e_B is sufficiently small requires measurements that are fast compared to memory decoherence times. Systems such as those based on ion traps can achieve this with good qubit memories^{12,19}.

Two additional assumptions are used. The first is that there is no error and no speed constraint on classical computations required to interpret measurement outcomes and control future gates. The second is that two-qubit gates can be applied to any pair of qubits without delay or additional error. This assumption is unrealistic, but the effect on the threshold is due primarily to CNOTs acting within the ancillas needed for maintaining one or two blocks encoding logical qubits. To account for this, we can use higher effective EPGs or require low-error quantum communication^{9,20,21}.

The above assumptions are standard in analyses of fault-tolerant architectures, but idealized. They are nevertheless believed to be sufficiently realistic that results based on them are meaningful in practice^{5,22–24}.

The quantum codes

The C_4/C_6 architecture is based on concatenating two quantum stabilizer codes, C_4 and C_6 . The codes are chosen to detect and correct errors with minimum effort. A stabilizer code is a common eigenspace of a set of commuting products of Pauli operators (the ‘check operators’). Such products are denoted by strings of X , Y , Z and I . For example, XIZ is a Pauli product on three qubits with X acting on the first and Z on the last. C_4 has check operators $XXXX$ and $ZZZZ$. It encodes a ‘qubit pair’ whose qubits may be labelled L and S , and defined by encoded operators $X_L = XXII$, $Z_L = ZIZI$, $X_S = IXIX$ and $Z_S = IIZZ$. C_4 is an optimal qubit-based one-error-detecting code. C_6 has check operators $XIIXXX$, $XXXIIX$, $ZIIZZZ$ and $ZZZIIZ$, which act on three consecutive qubit pairs. It encodes a qubit pair defined by encoded operators $X_L = IXXIII$, $Z_L = IIZZIZ$, $X_S = XIXXII$, $Z_S = IIIZZI$. C_6 is an optimal qubit-pair-based one-error-detecting code. The C_4/C_6 architecture uses C_4 to obtain level-1-encoded qubit pairs. We build subsequent levels by using three encoded qubit pairs to form a next-level C_6 -encoded qubit pair as shown in Fig. 1.

Given a joint eigenstate of the check operators, its list of eigenvalues is the ‘syndrome’. The level- l encoding has check operators that can be derived from the check and encoded operators of C_4 and C_6 . Ideally, the state of a level- l block has syndrome $\mathbf{0}$ (all eigenvalues are $+1$). In the presence of errors this is rarely the case, so the encoded qubits’ state is defined only with respect to a current ‘Pauli frame’ and an implicit recovery scheme. The Pauli frame is defined by a Pauli product that restores the error-free state of the block to the syndrome $\mathbf{0}$ subspace. The implicit recovery scheme determines the Pauli products needed to coherently map states with other syndromes to one with the error-free syndrome. By using the Pauli frame, we can avoid explicitly applying Pauli products for error correction and teleportation compensation^{9,25}. Error detection and correction are based on measurements that retroactively determine the syndrome of the state (the current syndrome has already been affected by further errors). An error is detected if the syndrome is not error-free according to the Pauli frame. In ‘postselected’ quantum computing, the state is then rejected and the computation restarted. In standard quantum computing, the syndrome

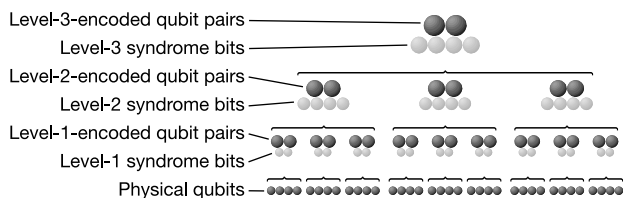


Figure 1 Block structure of C_4/C_6 concatenated codes. The bottom line shows nine blocks of four physical qubits. Each block encodes a level-1 qubit pair with C_4 . The encoded qubit pairs are shown in the line above. Formally, each such pair is associated with two syndrome bits, shown below the encoded pair in a lighter shade, which are accessible by syndrome measurements or decoding for error detection and correction. The next level groups three level-1 qubit pairs into a block, encoding a level-2 qubit pair with C_6 . The pair is associated with four syndrome bits. A level-2 block consists of 12 physical qubits. Three level-2 qubit pairs are used to form a level-3 qubit pair, again with C_6 and associated with four syndrome bits. The total number of physical qubits in a level-3 block is 36. In general, a level- l block has $4 \times 3^{l-1}$ physical qubits.

information is used to correct errors with a Pauli frame update. To do so, we use the fact that C_4 and C_6 can detect any error of one qubit and one qubit pair, respectively. If the location of the error is known, it can be corrected. This leads to the following error detection and correction (ED/EC) procedure: first we check the level-1 C_4 syndromes of each block of four qubits. For each block where an error is detected, mark the encoded level-1 qubit pair as having an error. Proceed to level 2 and check the (encoded) C_6 syndrome for each block of three level-1 pairs. If exactly one of the level-1 pairs has an error, use the C_6 syndrome to correct it. If not, mark the encoded level-2 pair as having an error unless none of the three level-1 pairs has an error and the C_6 syndrome is error-free according to the Pauli frame. Continue in this fashion through all higher levels. For optimizing state preparation, we can replace the error-correction step by error detection at the top few levels, depending on context, as explained below.

Error-correcting teleportation

To obtain syndrome information for a block B containing an encoded qubit pair we use error-correcting teleportation. We first prepare two blocks B_1 and B_2 , each encoding a logical qubit pair so that the first pair is maximally entangled with the second, in the logical state $(|0000\rangle + |0101\rangle + |1010\rangle + |1111\rangle)/2$. B_1 and B_2 form an ‘encoded (or logical) Bell pair’. The encoded Bell pair is prepared ‘fault-tolerantly’, so that each block’s errors are essentially as though the physical qubits were subject to independent errors of order γ . The next step is to apply Bell measurements (the first step of conventional quantum teleportation²⁶) to corresponding physical qubits in B and B_1 . This results in the transfer of B ’s encoded state to B_2 , up to a known change in the Pauli frame. It can be shown that the Bell measurement outcomes reveal the eigenvalues of the products of corresponding check operators on B and B_1 , which is sufficient for inferring the needed syndrome for error detection and correction. Error detection or correction is successful if the combined errors from B and B_1 are within the capabilities of the codes. See ref. 16 for further details.

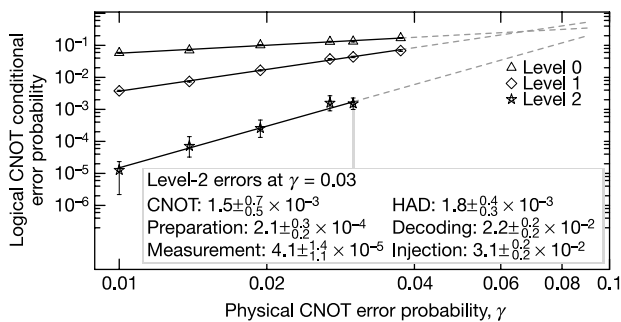


Figure 2 Conditional logical errors with postselection. The plot shows logical CNOT errors conditional on not detecting errors as a function of EPG parameter γ at levels 0, 1 and 2. The logical CNOT is implemented with transversal physical CNOTs and two error-detecting teleportations, where the output state is accepted only if no errors are detected in the teleportations. The data show the incremental error attributable to the logical CNOT in the context of a longer computation (Supplementary Information B). The error bars are 68% confidence intervals. The solid lines are obtained by gradient-descent likelihood maximization. Extrapolations are shown with dashed lines and suggest that logical EPG improvements with increasing levels are possible above $\gamma = 6\%$. Other operations’ errors for $\gamma = 3\%$ and level 2 are shown in the inset table. The decoding error is the incremental error introduced by decoding a block into two physical qubits. The injection error is the error in a logical state that we prepare by decoding one block of a logical Bell pair and measuring the decoded qubits. Decoding and injection errors were found to decrease from level 1 (decoding error $4.4 \pm 0.4 \times 10^{-2}$, injection error $5.5 \pm 0.5 \times 10^{-2}$) to level 2.

Encoded state preparation

Fault-tolerant architectures depend on having a plentiful supply of verified, fault-tolerantly prepared encoded states. In the C_4/C_6 architecture, encoded Bell pairs are fundamental to preparing all other such states. An encoded Bell pair on blocks B_1 and B_2 is prepared at level $l + 1$ from level- l -encoded Bell pairs in three steps. The first step is to prepare level- $l + 1$ -encoded $|00\rangle$ in block B_1 and $|++\rangle$ in block B_2 . C_4 and C_6 have the property that these states are local variants of level- l ‘cat’ states (states such as $(|0\dots 0\rangle + |1\dots 1\rangle)/\sqrt{2}$), which can be obtained and verified by linking level- l Bell pairs. The second step is a ‘transversal’ CNOT consisting of physical CNOTs applied from qubits of B_1 to corresponding qubits of B_2 . The third step involves error-correcting teleportations of level- l sub-blocks of B_1 and B_2 , which is required to manage errors introduced in the first two steps and limit correlations between B_1 and B_2 . The state preparation networks are shown in Supplementary Information A.

Logical Clifford gates for C_4 and C_6

For simplicity, we treat the qubits in a logical qubit pair identically and ignore one of them for the purpose of computation. Preparation of logical $|00\rangle$ and $|++\rangle$ is accomplished by using the

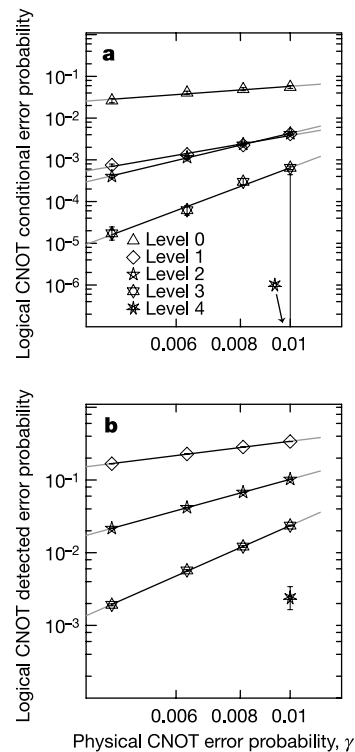


Figure 3 Conditional and detected logical errors with error correction. The plot shows incremental detected and conditional logical errors for a logical CNOT as a function of EPG parameter γ up to level 4. Error bars and lines are as described in the caption of Fig. 2. The combination of error correction and detection is as required for the error-correcting C_4/C_6 architecture. **a**, The logical CNOT’s error conditional on not detecting an uncorrectable error. **b**, The probability of detecting an uncorrectable error. At $\gamma = 1\%$, the detected errors are $2.4 \pm 0.0 \times 10^{-2}$ (level 3) and $2.4 \pm 0.7 \times 10^{-3}$ (level 4). The conditional errors are $6.4 \pm 0.6 \times 10^{-4}$ (level 3) and $0.0 \pm 0.0 \times 10^{-4}$ (level 4). For comparison, the preparation errors at levels 3 and 4, respectively were found to be $2.1 \pm 0.3 \times 10^{-4}$ and $0.0 \pm 1.0 \times 10^{-4}$ (detected errors) and $3.3 \pm 7.5 \times 10^{-6}$ and $0.0 \pm 1.0 \times 10^{-4}$ (conditional errors). The measurement errors are $4.7 \pm 0.4 \times 10^{-4}$ and $5.6 \pm 12.8 \times 10^{-5}$ (detected errors) and $3.3 \pm 2.7 \times 10^{-6}$ and $0.0 \pm 1.0 \times 10^{-4}$ (conditional errors). Finally, the HAD errors at level 3 are $1.3 \pm 0.0 \times 10^{-2}$ (detected error) and $3.5 \pm 0.5 \times 10^{-4}$ (conditional error).

first step of the logical Bell pair preparation procedure followed immediately by error-correcting teleportations of the sub-blocks. The codes C_4 , C_6 and their concatenations have the property that logical CNOTs and measurements can be implemented transversally²². This ensures fault tolerance. The HAD gate can be implemented transversally with a permutation of the physical qubits in a block. Permutations can be implemented by relabelling without physical manipulations and are also fault-tolerant. To control error propagation, we include with each logical gate error-correcting teleportations of its blocks.

$\mathcal{G}_{\min}$ thresholds

For the purpose of establishing high thresholds, we first consider postselected $\mathcal{G}_{\min}$ computing. Postselected computing is a model of computing that abstracts and generalizes the key non-deterministic aspects of techniques such as purification²⁷ and verified state preparation²⁸. Here we use it to prepare states needed for scalable quantum computing without having to specify the state-preparation networks. Postselected computing is like standard quantum computing except that when a gate is applied, the gate may fail. If it fails, this is known. The probability of success must be non-zero. There may be gate errors conditional on success, but fault-tolerant postselected computing requires that such errors are small. We implement fault-tolerant postselected computing with the C_4/C_6 architecture by aborting the computation whenever an error is detected. Error-correcting teleportation is replaced by error-detecting teleportation, which uses the syndrome information only for error detection. In ref. 23 we used a computer-assisted heuristic analysis to obtain a threshold value of 3%, below which fault-tolerant postselected $\mathcal{G}_{\min}$ computing is possible. Here we use direct simulation of the error behaviour of postselected encoded CNOTs with error-detecting teleportation at up to two levels of encoding and physical EPGs of $1\% \leq \gamma \leq 3.75\%$. The simulation method is explained in Supplementary Information B. The simulated conditional logical errors are shown in Fig. 2 and suggest a threshold of above 6% by extrapolation.

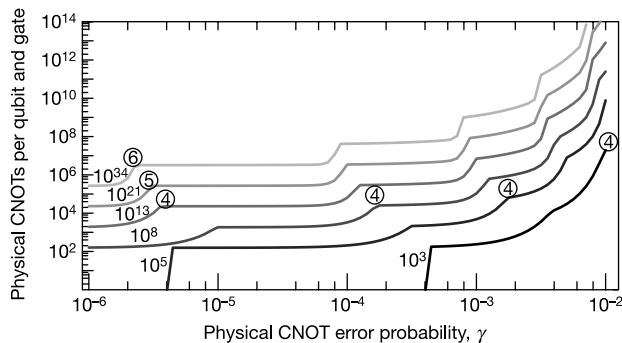


Figure 4 Estimating C_4/C_6 resource requirements. The figure shows the number N_{CNOT} of physical CNOTs required per qubit and gate to implement computations of sizes $G = 10^3, 10^5, \dots, 10^{34}$ (curves with G indicated). Other resources are dominated by N_{CNOT} . An order-of-magnitude estimate of the total number of physical CNOTs required by a computation can be made as follows: we determine the number G of gates required, including ‘memory’ gates. Using the corresponding curve in the figure, we find N_{CNOT} at the physical EPG. We then multiply N_{CNOT} by G and the average number of additional logical CNOTs per gate required for fault-tolerantly preparing and purifying states such as logical $|\pi/8\rangle$. A conservative estimate for the latter number is 300. With maximum parallelism, the ‘scale-up’ (number of physical qubits per logical qubit) is of the same order as N_{CNOT} . If the memory error is not too large, this can be reduced to about $2(1 + 2(-1))3^{-1}$ with moderate parallelism. The circled numbers are at the points on each curve above which the indicated level of concatenation must be used. Levels increment at each step-like feature of the curves.

Scalable $\mathcal{G}_{\min}$ computing with the C_4/C_6 architecture requires lower EPGs and the use of error correction to increase the probability of success to near 1. To optimize the resource requirements needed to achieve a given logical EPG, the last level at which error correction is used in the ED/EC procedure is d_l levels below the relevant top level, where d_l depends on context and γ . At higher levels, errors are only detected. For simplicity and to enable extrapolation by modelling, we examined a fixed strategy with $d_l = 1$ in all state-preparation contexts and $d_l = 0$ (maximum error correction) in the context of logical computation. The relevant top level in a state preparation context is the level of a block measurement or error-correcting teleportation of a sub-block, not the logical level of the state that is eventually prepared. Each logical gate now has a probability of detected but uncorrectable error, and a probability of logical error conditional on not having detected an error. Figure 3 shows both error probabilities up to level 4 for a logical CNOT with error-correcting teleportation and EPGs of $\gamma \leq 1\%$. The data indicate that the $\mathcal{G}_{\min}$ threshold for this architecture is above 1%.

Universal computation

To complete the $1\mathcal{G}_{\min}$ gate set so that we can implement arbitrary quantum computations, it suffices to add HAD and preparation of the state $|\pi/8\rangle = \cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle$ ^{29,30} in both qubits of a logical pair. The logical errors of HAD are less than those of the logical CNOT. To prepare logical $|\pi/8\rangle|\pi/8\rangle$ in a qubit pair, we obtain a logical Bell pair, decode its first block into two physical qubits and make measurements to project the physical qubits’ states onto $|\pi/8\rangle$ or the orthogonal state. If an orthogonal state is obtained, we adjust the Pauli frame by Y operators accordingly. Because of the entanglement between the physical qubits and the logical ones, this prepares the desired logical state, albeit with error. This procedure is ‘state injection’. To decode the first block of the Bell pair, we first decode the C_4 sub-blocks and continue by decoding six-qubit sub-blocks of C_6 . Syndrome information is obtained in each step and can be used for error detection or correction. The error in decoding is expected to be dominated by the last decoding steps. Consequently, the error in the injected state should be bounded as the number of levels increase, which we verified by simulation. To remove errors from the injected states, logical purification can be used^{30,31} and is effective if the error of the injected state is less than 0.141 (ref. 31). The purification method can be implemented fault-tolerantly to ensure that the purified logical $|\pi/8\rangle$ states have errors similar to those of logical CNOTs (Supplementary Information C).

Consider the threshold for fault-tolerant postselected universal quantum computing. The logical HAD and injection errors at $\gamma = 3\%$ and level 2 are shown in Fig. 2. The injection error is well below the maximum allowed and is not expected to increase substantially for higher levels. The injection error should scale approximately linearly with EPG, so the extrapolated threshold above 6% may apply.

The injection and purification method for preparing states needed to complete the gate set works with the error-correcting C_4/C_6 architecture. Consider state injection at $\gamma = 1\%$. The context for injection is state preparation, which determines the combination of error correction and detection as discussed above. The conditional logical error after state injection was determined to be $8.6 \pm_{0.5}^{0.6} \times 10^{-3}$ at level 3 and $1.1 \pm_{0.1}^{0.1} \times 10^{-2}$ at level 4, comparable to γ and sufficiently low for $|\pi/8\rangle$ purification. As a result, the C_4/C_6 architecture enables scalable quantum computing at EPGs above 1%.

To obtain higher thresholds, we use fault-tolerant postselected computing to prepare states in a code that can handle higher EPGs than C_4/C_6 concatenated codes can. The states are chosen so that we can implement a universal set of gates by error-correcting teleportation. Suppose that arbitrarily low logical EPGs are achievable with

the C_4/C_6 architecture for postselected computing. To compute scalably, we choose a sufficiently high level l for the C_4/C_6 architecture and a very good error-correcting quantum code C_e . The first step is to prepare the desired C_e -encoded states using level- l -encoded qubits, in essence concatenating C_e onto level l of the C_4/C_6 architecture. Suppose that the conditional error in the logical state prepared can be made arbitrarily small. The second step is to decode each block of the C_4/C_6 architecture to physical qubits to obtain unconcatenated C_e -logical states (partial decoding). Once these states are successfully prepared, we use them to implement each logical gate by error-correcting teleportation. Simulations show that the postselected decoding introduces an error $\leq \gamma$ for each decoded qubit (Fig. 2). There is no postselection in error-correcting teleportation with C_e , and it is sensitive to decoding error in two blocks ($\sim 2\gamma$) as well as the error of the CNOT ($\sim \gamma$) and the two physical measurements ($\sim 8\gamma/15$) required for the Bell measurement. Hence, the effective error per qubit that needs to be corrected is $\sim 3.53\gamma$. The maximum error probability per qubit correctable by known codes C_e is ~ 0.19 (ref. 32). Thus, if $\gamma \leq 5\%$ (conservatively below $0.19/3.53$) the C_e architecture can have small logical errors, say below 10^{-3} . Scalable quantum computing is then possible by using C_e -encoded qubits as the founding qubits for the error-correcting C_4/C_6 architecture (for example). Because 5% is below the extrapolated threshold for the postselected C_4/C_6 architecture, scalable quantum computing may be possible with our architecture at EPGs as high as 5% (or at least 3% without extrapolation). Although the postselection overheads are extreme, the above architecture is theoretically efficient: the asymptotic overheads for implementing a quantum computation are polynomial in terms of the computation's size.

Resources

The resource requirements for the error-correcting C_4/C_6 architecture can be mapped out as a function of γ for different sizes of computations. We do not have analytical expressions for the resources for logical Bell pair preparation or for the logical errors as a function of γ and, with our current capabilities, we are not able to determine them in enough detail by simulation. We therefore use naive models to approximate the expressions needed. The number of physical CNOTs used in a logical Bell-pair preparation is modelled by functions of the form $C/(1-\gamma)^k$, which would be correct on average if the state-preparation network had C gates of which k failed independently with probability γ , and the network was repeatedly applied until none of the k gates failed. C and k depend on the level of concatenation. The logical error probabilities are modelled at level $l \geq 1$ by $p_d(l) = d(l)\gamma^{f(l+1)}$ (detected error) and $p_c(l) = c(l)\gamma^{f(l+2)}$ (conditional logical error), where $f(0) = 0$, $f(1) = 1$, $f(l+1) = f(l) + f(l-1)$ is the Fibonacci sequence. These expressions are asymptotically correct as $\gamma \rightarrow 0$. We verified that they model the desired values well and determined the constants at low levels by simulation and at high levels by extrapolation (Supplementary Information C).

Figure 4 shows the resource requirements as a function of computation size. Following the instructions in the caption, we obtain the following order-of-magnitude estimates: at an EPG of 1%, a computation with 10^3 or 10^5 gates and (say) 100 or more qubits requires 6×10^{12} or 2×10^{17} physical CNOTs, respectively. A more precise calculation shows that 1.2×10^{14} physical CNOTs are required for 1,000 logical $|\pi/8\rangle$ preparations (Supplementary Information C). Given current capabilities, the outputs of these computations are not predictable with classical algorithms. The quantum resource requirements are large and at present difficult to realize. However, comparable complexity is achieved in today's classical computers: central processing units have 10^8 or more transistors operating at rates of 10^9 steps per second³³, making available up to 10^{17} bit operations per second.

The resource requirements decrease rapidly with lower EPGs. At EPGs well below 10^{-3} , an architecture based on unconcatenated block codes such as that of Steane⁹ is expected to be more efficient. Indeed, at an EPG of 10^{-4} , such architectures use one to two orders of magnitude fewer resources. The C_4/C_6 architecture still has the advantage of simplicity, and of yielding more reliable answers conditional on having no detected errors.

Discussion

An important use of studies of fault-tolerant architectures is to provide guidelines for EPGs that should be achieved to meet the low-error criterion for scalability. Such guidelines depend on the details of the relevant error models and constraints on two-qubit gates. Nevertheless, the value of $\gamma = 10^{-4}$ has often been cited as the EPG to be achieved. With architectures such as that of Steane^{9,34} and the one introduced here, resource requirements at $\gamma = 10^{-3}$ are now comparable to what they were for $\gamma = 10^{-4}$ at the time this value was starting to be cited²².

Several open problems arise from the work presented here. Can the high thresholds evidenced by our simulations be mathematically proved? Are thresholds for postselected computing strictly higher than thresholds for scalable standard quantum computing? Recent work by Reichardt³⁴ shows that Steane's architecture can be made more efficient by the judicious use of error detection, improving Steane's threshold estimates to around 10^{-2} . How do the available fault-tolerant architectures compare for EPGs between 10^{-3} and 10^{-2} ? It would be helpful to improve significantly the resource requirements of fault-tolerant architectures, particularly at high EPGs. □

Received 12 November 2004; accepted 4 January 2005; doi:10.1038/nature03350.

- Shor, P. W. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **26**, 1484–1509 (1997).
- Feynman, R. P. Simulating physics with computers. *Int. J. Theor. Phys.* **21**, 467–488 (1982).
- Abrams, D. S. & Williams, C. P. Fast quantum algorithms for numerical integrals and stochastic processes. Preprint at (<http://arXiv.org/quant-ph/9908083>) (1999).
- DiVincenzo, D. The physical implementation of quantum computation. *Fort. Phys.* **48**, 771–783 (2000).
- Preskill, J. Reliable quantum computers. *Proc. R. Soc. Lond. A* **454**, 385–410 (1998).
- Kitaev, A. Y. Quantum computations: Algorithms and error correction. *Russ. Math. Surv.* **52**, 1191–1249 (1997).
- Aharonov, D. & Ben-Or, M. Fault-tolerant quantum computation with constant error. Preprint at (<http://www.arXiv.org/quant-ph/9906129>) (1999).
- Knill, E., Laflamme, R. & Zurek, W. H. Resilient quantum computation. *Science* **279**, 342–345 (1998).
- Steane, A. M. Overhead and noise threshold of fault-tolerant quantum error correction. *Phys. Rev. A* **68**, 042322 (2003).
- Gottesman, D. *Stabilizer Codes and Quantum Error Correction*. PhD thesis, California Institute of Technology, Pasadena (1997).
- Leibfried, D. et al. Experimental demonstration of a robust, high-fidelity geometric two ion-qubit phase gate. *Nature* **422**, 412–415 (2003).
- Roos, C. F. et al. Bell states of atoms with ultralong lifetimes and their tomographic state analysis. *Phys. Rev. Lett.* **220402** (2004).
- Knill, E., Laflamme, R., Martinez, R. & Tseng, C.-H. An algorithmic benchmark for quantum information processing. *Nature* **404**, 368–370 (2000).
- Childs, A. M., Chuang, I. L. & Leung, D. W. Realization of quantum process tomography in NMR. *Phys. Rev. A* **64**, 012314 (2001).
- Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, Cambridge, UK, 2001).
- Knill, E. Scalable quantum computation in the presence of large detected-error rates. Preprint at (<http://www.arXiv.org/quant-ph/0312190>) (2003).
- Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
- Zhou, X., Leung, D. W. & Chuang, I. L. Methodology for quantum logic gate construction. *Phys. Rev. A* **62**, 052316 (2000).
- Bollinger, J. J., Heinzen, D. J., Itano, W. M., Gilbert, S. L. & Wineland, D. J. A 303 MHz frequency standard based on trapped Be^+ ions. *IEEE Trans. Instrum. Meas.* **40**, 126–128 (1991).
- Steane, A. M. Quantum computer architecture for fast entropy extraction. *Quant. Inf. Comput.* **4**, 297–306 (2002).
- Svore, K. M., Terhal, B. M. & DiVincenzo, D. P. Local fault-tolerant quantum computation. Preprint at (<http://www.arXiv.org/quant-ph/0410047>) (2004).
- Steane, A. Space, time, parallelism and noise requirements for reliable quantum computing. *Fort. Phys.* **46**, 443–457 (1998).
- Knill, E. Fault-tolerant postselected quantum computation: Threshold analysis. Preprint at (<http://www.arXiv.org/quant-ph/0404104>) (2004).
- Knill, E. Quantum computing with very noisy devices. Preprint at (<http://www.arXiv.org/quant-ph/0410199>) (2004).

25. Raussendorf, R. & Briegel, H. J. A one-way quantum computer. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
26. Bennett, C. H. *et al.* Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
27. Bennett, C. H. *et al.* Purification of noisy entanglement and faithful teleportation via noisy channels. *Phys. Rev. Lett.* **76**, 722–725 (1996).
28. Shor, P. W. In *Proc. 37th Symp. Foundations of Computer Science (FOCS)* 56–65 (IEEE Press, Los Alamitos, California, 1996).
29. Knill, E., Laflamme, R. & Zurek, W. Resilient quantum computation: Error models and thresholds. *Proc. R. Soc. Lond. A* **454**, 365–384 (1998).
30. Knill, E. Fault-tolerant postselected quantum computation: Schemes. Preprint at (<http://www.arXiv.org/quant-ph/0402171>) (2004).
31. Bravyi, S. & Kitaev, A. Universal quantum computation based on a magic states distillation. Preprint at (<http://www.arXiv.org/quant-ph/0403025>) (2004).
32. DiVincenzo, D. P., Shor, P. W. & Smolin, J. A. Quantum-channel capacity of very noisy channels. *Phys. Rev. A* **57**, 830–839 (1998).
33. Intel Cooperation. Microprocessor quick reference guide. (<http://www.intel.com/pressroom/kits/quickreffam.htm>) (2004).
34. Reichardt, B. W. Improved ancilla preparation scheme increases fault tolerant threshold. Preprint at (<http://www.arXiv.org/quant-ph/0406025>) (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work is a contribution of NIST, an agency of the US government, and is not subject to US copyright. Partial support from the DARPA QuIST programme is acknowledged.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to E.K. (knill@boulder.nist.gov).

A biodiversity intactness index

R. J. Scholes & R. Biggs

CSIR Environmentek, PO Box 395, Pretoria, 0001, South Africa

The nations of the world have set themselves a target of reducing the rate of biodiversity loss by 2010. Here, we propose a biodiversity intactness index (BII) for assessing progress towards this target that is simple and practical—but sensitive to important factors that influence biodiversity status—and which satisfies the criteria for policy relevance set by the Convention on Biological Diversity. Application of the BII is demonstrated on a large region ($4 \times 10^6 \text{ km}^2$) of southern Africa. The BII score in the year 2000 is about 84%: in other words, averaged across all plant and vertebrate species in the region, populations have declined to 84% of their presumed pre-modern levels. The taxonomic group with the greatest loss is mammals, at 71% of pre-modern levels, and the ecosystem type with the greatest loss is grassland, with 74% of its former populations remaining. During the 1990s, a population decline of 0.8% is estimated to have occurred.

The loss of biodiversity in the modern era, at rates unequalled since the major extinction events in the distant geological past^{1,2}, is a matter of considerable policy concern. The Convention on Biological Diversity (CBD) has adopted a target of reducing the rate of biodiversity loss by 2010 (ref. 3). For this target to be met, a method of measuring biodiversity status must be agreed on and implemented. At present no scientific consensus measure exists⁴, although several candidates have been proposed⁵. The CBD has agreed on a partial set of indicators⁶. Difficulties in establishing operational indicators stem largely from the complex, multi-dimensional nature of biodiversity, which can be defined in terms of composition, structure and function at multiple scales⁷.

The CBD developed a set of criteria that an indicator of biodiversity change should satisfy⁸. The salient ones are that it should be scientifically sound, be sensitive to changes at policy-relevant spatial and temporal scales, allow for comparison with a baseline situation and policy target, be useable in models of future projections, and be amenable to aggregation and disaggregation at ecosystem, national and international levels. It should also be simple and easily understood, broadly accepted and measurable with sufficient accuracy at affordable cost.

Whereas there is no shortage of ways to express biodiversity^{8,9}, none adequately meets the criteria set by the CBD. Most indices require essentially complete knowledge of the biota or the population sizes of individual species, neither of which are achievable conditions at regional to global scales for the next several decades. Many methods are scale-dependent and thus hard to interpret in a comparative context. The most widely used indicators are based either on risk of extinction¹⁰, or on land area under conservation protection¹¹. Several indices have been offered that combine either a sparse and selective set of population estimates for indicator species¹², or combine a number of factors that are thought to relate to biodiversity status¹³.

A sensitive, realistic and useful measure of biodiversity loss needs to be based on changes in population abundance across a wide range of species, and must consider the entire landscape. At a global scale, habitat loss, including reductions in both quality and quantity of suitable environment, is the main factor responsible for declines in species abundance^{1,14}. Other important causes, such as excessive harvest pressure or the effects of pollutants, can also be expressed on the basis of area affected and intensity of impact. This paper introduces an index that takes into account these factors and meets the CBD criteria for policy relevance.

Attributes of the proposed index

The BII is an indicator of the average abundance of a large and diverse set of organisms in a given geographical area, relative to their

reference populations. We recommend calculating the BII across all species within the broad taxonomic groups that are reasonably well described. For most parts of the world this includes plants and vertebrates, and excludes invertebrates and microbes, which are diverse but poorly documented. We exclude alien species in the calculation of the index.

An easily grasped reference population for large parts of the world is that which occurred in the landscape before alteration by modern industrial society. Because accurate data on pre-modern populations are seldom available, contemporary populations in large protected areas serve as a practical reference. Alternative reference points can be defined for parts of the world that had already been highly transformed by the beginning of the modern period; for example, by selecting a specific baseline year within records or reliable memory.

The BII can in principle be calculated exactly by 'bottom-up' aggregation of population data for individual species. However, especially in the highly biodiverse, but poorly studied parts of the world, this will not be a practical option for the next several decades. The proposed strategy is therefore to initially calculate the BII 'top-down'. This is analogous to the top-down/bottom-up distinction that has permitted progress in the assessment of global greenhouse gas emissions¹⁵. In the greenhouse gas example, the collective national emissions from thousands of individual sources are estimated on an activity basis, rather than source-by-source summation. In the case of biodiversity, we estimate the impacts of a set of land use activities on the population sizes of groups of ecologically similar species ('functional types'). The chosen land use activities range from complete protection to extreme transformation, such as urbanization. All activities are expressed on the basis of the area affected. The index is aggregated by weighting by the area subject to each activity and the number of species occurring in the particular area.

The BII is an aggregate index, intended to provide an intuitive, high-level synthetic overview for the public and policy makers. It can be disaggregated in several ways to meet the information needs of particular users: by ecosystem or political units, taxonomic group, functional type, or land use activity. This provides transparency and credibility. The BII has the same meaning at all spatial scales. It is possible to estimate the value of BII for the past, and project it into the future under various situations. An error bar can be associated with the BII, allowing a monitoring goal to be defined in terms of shrinking the uncertainty range.

The biodiversity intactness index algorithm

The BII gives the average richness- and area-weighted impact of a set of activities on the populations of a given group of organisms in a

specific area. The population impact (I_{ijk}) is defined as the population of species group i under land use activity k in ecosystem j , relative to a reference population in the same ecosystem type. The BII is calculated as:

$$BII = (\sum_i \sum_j \sum_k R_{ij} A_{jk} I_{ijk}) / (\sum_i \sum_j \sum_k R_{ij} A_{jk})$$

where R_{ij} = richness (number of species) of taxon i in ecosystem j , and A_{jk} = area of land use k in ecosystem j

The BII can be disaggregated along different axes. For instance, the intactness of a particular taxonomic group i is given by:

$$BII_i = (\sum_j \sum_k R_{ij} A_{jk} I_{ijk}) / (\sum_j \sum_k R_{ij} A_{jk})$$

For ease of understanding, we suggest expressing BII as a percentage rather than a proportion.

Species-by-species population data for estimating I_{ijk} are seldom available for more than a few species, in a few locations. We used expert judgement to generate a matrix of values of I_{ijk} for southern Africa. Three or more highly experienced specialists in each broad taxonomic group (plants, mammals, birds, reptiles and amphibians) were independently asked to estimate the reduction in the populations of their speciality group caused by a predefined set of land use activities (Table 1). Their estimates were made relative to populations in a large protected area in the same ecosystem type, of which six were defined: forest, savanna, grassland, shrubland, *fynbos* (a South African sclerophyllous thicket) and wetland. To assist in the estimation process, each taxonomic group was divided into five to ten functional types that respond in similar ways to human activities. Across all groups, the functional types were defined primarily by body size, trophic niche and reproductive strategy. I_{ijk} typically assumed values between 0% and 100%, but exceeded 100% in situations where certain activities benefited particular functional types. Estimates of I_{ijk} were aggregated up to the broad taxonomic level by weighting the estimates for each functional type by the number of species in that group in the particular ecosystem type.

A total of 4,650 estimates of I_{ijk} were made, comprising five broad

taxonomic groups, each with at least three experts, six ecosystem types, an average of six land use activities and eight functional types. This works out at approximately 300 estimates per expert, a process that took about five hours per interview. Estimating the BII for southern Africa therefore took a few weeks of effort, rather than the decades needed for detailed population surveys. The range of estimates from different experts was used to construct an uncertainty bar around I_{ijk} (Fig. 1). Expert-derived estimates of I_{ijk} were validated against measurements available in the literature^{16–26} (Fig. 2). The paucity of field studies, and the fact that the variation between comparable field studies is substantially larger than between the expert estimates, supports the use of expert-based approaches at this time.

Species richness data (R_{ij}) is typically available as total species counts per ecosystem type. In this study, species richness data²⁷ associated with the WWF ecoregions²⁸ were used. Using such data is equivalent to assuming that every species occurs throughout the extent of the particular ecosystem type. The BII can also be calculated using the potential geographical distributions for individual species, where such data are available.

The area of a particular land use within a specific ecosystem type, A_{jk} , is determined by overlaying land use and ecosystem maps. In this study, broad classes of land use were inferred from land cover and land tenure boundaries. We suggest limiting the number of land use classes to below ten, to keep the number of I_{ijk} estimates required manageable. We defined and mapped six levels of land use intensity (Table 1). Where classes derived from different data sources overlapped, the highest impact land use was assigned. The resolution of the land use classification affects the estimation of I_{ijk} . In regional-to-global studies, land cover is typically mapped at a resolution of about 1×1 km, and an area classified as, for example, 'cultivated' almost always has inclusions of uncultivated land. Experts were instructed to take this into account in making their estimates of I_{ijk} .

How finely the taxonomic groups need to be divided into functional types, the broad biomes into particular ecosystem

Table 1 **Classes and data sources used to compile a land use map**

Land use class	Description	Examples	Data source
Protected	Minimal recent human impact on structure, composition or function of the ecosystem. Biotic populations inferred to be near their potential.	Large protected areas, national, provincial and private nature reserves, 'wilderness' areas.	World Database on Protected Areas ¹¹ . All designated protected areas of IUCN categories I–V.
Moderate use	Extractive use of populations and associated disturbance, but not enough to cause continuing or irreversible declines in populations. Processes, communities and populations largely intact.	Forest areas used by indigenous peoples or under sustainable, low-impact forestry; grasslands grazed within their sustainable carrying capacity.	All remaining areas not classified into one of the other five categories.
Degraded	Extractive use at a rate exceeding replenishment and widespread disturbance. Often associated with high human population densities and poverty in rural areas. Productive capacity reduced to approximately 60% of 'natural' state.	Clear-cut logging, areas subject to intense harvesting, hunting, fishing or overgrazing, areas invaded by alien vegetation.	All areas falling below 75% (forest, grassland and savanna) or 50% (shrublands) of expected production as estimated by nonlinear regression (Michaelis–Menten function) of maximum annual NDVI on growth days. Degraded areas not estimated for desert, wetland and <i>fynbos</i> .
Cultivated	Natural land cover replaced by planted crops. Most processes persist, but are significantly disrupted by ploughing and harvesting activities. Residual biodiversity persists in the landscape, mainly in set-asides and in strips between fields (matrix), assumed to constitute approximately 20% of class.	Commercial and subsistence crop agriculture, both irrigated and dryland, including planted pastures and fallow, or recently abandoned cultivated areas. Orchards and vineyards.	SADC Landcover Data set ³⁶ , filled with GLC2000 (ref. 37) for Namibia and Botswana.
Plantation	Natural land cover permanently replaced by dense plantations of trees. Unplanted areas assumed to constitute approximately 25% of class.	Plantation forestry, typically <i>Pinus</i> and <i>Eucalyptus</i> species.	SADC Landcover Data set ³⁶ .
Urban	Land cover replaced by hard surfaces such as roads and buildings. Dense populations of people. Most ecological processes are highly modified. Remnant semi-natural cover assumed to constitute 10% of class.	Dense human settlements, industrial areas, transport infrastructure, mines and quarries.	Urban extents ³⁸ .

types, and how many land use activities are defined is largely a pragmatic question. Our experience is that the patterns of impact of a given land use activity is markedly similar between ecosystem types (confirmed by analysis of variance (ANOVA), which showed

no significant biome by land use interaction). This suggests that defining one set of impacts per functional type per biome (high-level ecosystem classification) is sufficient, despite the fact that even the relatively coarse-level ecoregion classification we used for species richness²⁸ has several ecosystem types per biome.

The uncertainty in BII can be calculated from the standard errors of each component in the BII equation. We found that by far the largest error was associated with I_{ijk} . We used this error to calculate a 95% confidence interval for BII. We tested the sensitivity of BII to different resolutions and sources of data for A_{jk} and R_{ij} and found it to be small in the southern African case.

The biodiversity intactness of southern Africa

We applied the BII to the region of southern Africa comprising South Africa, Namibia, Lesotho, Swaziland, Botswana, Zimbabwe and Mozambique. This region was selected because the experts we interviewed felt comfortable extrapolating their experience within this biogeographical domain, but not beyond. Overall, we estimate that $84 \pm 7\%$ of the pre-colonial number of wild organisms persist in present-day southern Africa (Table 2), despite greatly increased human demands on ecosystems over the past 300 years. In contrast, over 99% of the species persist¹, illustrating the insensitivity of indices based on extinction (changes in richness alone). Protected areas currently constitute 8% of the region, moderate use areas 78%, degraded areas 2%, cultivated areas 9%, plantations 0.5% and urban areas 2% (Table 1). The persistence of populations of wild organisms is therefore substantially greater than that indicated by the area under formal conservation, and points to the importance of areas outside of reserves in the preservation of biodiversity. The contemporary distribution of wild organisms in southern Africa is 85% in moderate use areas, 10% in protected areas, and the remaining 5% in cultivated, degraded, urban and plantation areas.

The impact of humans on biodiversity is expressed very selectively. Large-bodied mammal, bird and reptile species that are easy to hunt or harvest are most affected, especially if they are valuable or in direct conflict with human well-being. These species are only a small portion of the total biodiversity. The vast majority of species are affected mainly through loss of habitat to cultivation or urban settlement, both of which make up relatively small fractions of the southern African landscape.

The greatest impact on biodiversity in southern Africa has occurred in the grassland biome, followed closely by *fynbos*. In both cases, the main cause is conversion to cultivated land, followed

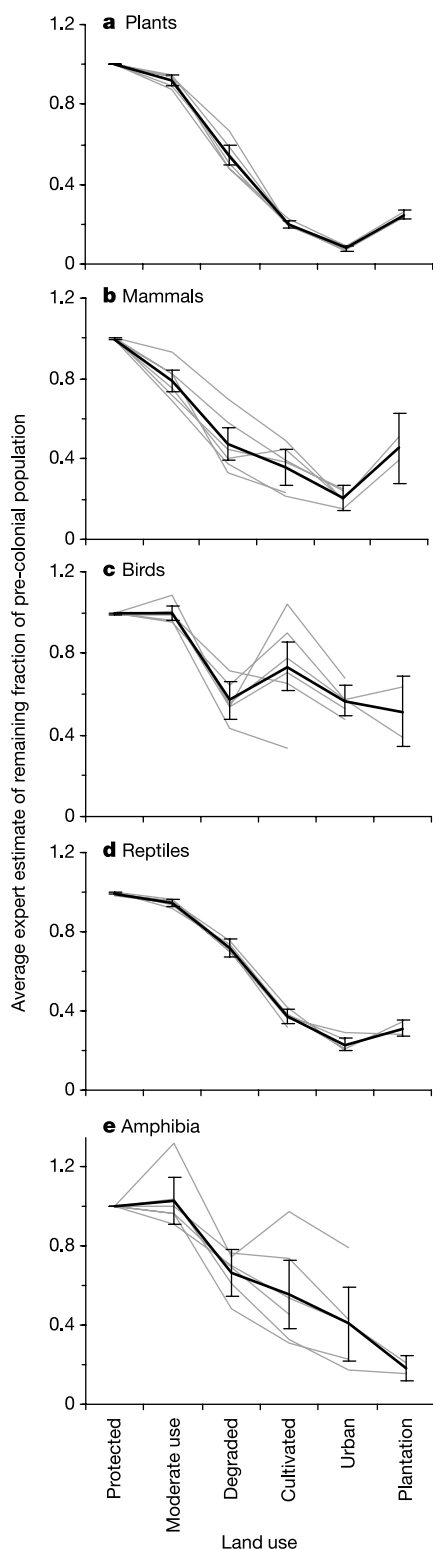


Figure 1 Average fraction of original populations remaining under a range of land use activities. Data are estimates by three or more taxon experts for each broad taxonomic group (a–e). The grey lines reflect the average estimates for each of six biomes or ecosystem types and the bold line reflects the estimated impact averaged across all biomes. The error bar reflects the 95% confidence interval around the mean estimate.

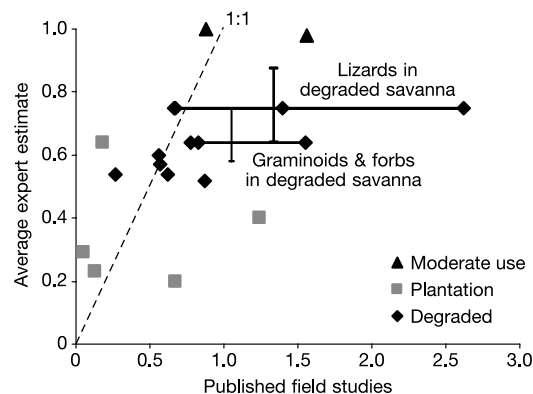


Figure 2 Comparison of estimates of I_{ijk} , the ratio of current populations to reference populations under various land use activities. This ratio was determined by expert estimation versus field studies published in the literature ($n = 19$, $r = 0.47$, $P = 0.04$). For the two functional groups (lizards and graminoids and forbs) where several comparable field studies were available, the range (bar) reported by field studies was substantially larger than the range of estimates between experts.

by urban sprawl and plantation forestry. The arid shrublands and savannas that make up most of the land area of southern Africa are overall less affected. The main cause of biodiversity loss in these systems is land degradation, defined here as land uses that do not alter the cover type, but lead to a persistent loss in ecosystem productivity. Aggregated at a national level, the most densely populated countries in the region, Lesotho and Swaziland, have lower BII scores than the sparsely inhabited countries of Botswana and Namibia (Table 3).

We estimated the rate of change of BII in southern Africa during the decade of the 1990s as an absolute decline of 0.8%. We used data on average rates of land cover change in the region for this calculation. Over the period 1991 to 2000, the relative increase in protected areas was 3.3% (ref. 11), cultivated areas 10.1% (ref. 29), urban areas 5.1% (ref. 29), and plantation forestry 9.2% (ref. 30). Degradation was assumed to have kept pace with the growth of the rural population²⁹ (12.9%), and the moderate use category was calculated by difference (−1.9%). Assuming that the I_{ijk} values remained constant over the decade, we calculated the overall BII for the year 1990 as 85.2%.

Our results suggest that the policy action with the greatest potential to prevent further loss of biodiversity in southern Africa is to prevent the extensive areas currently under moderate extractive use from becoming degraded. Moderately used land (for example, grazed within stocking norms) has almost the same level of biodiversity as protected areas. Degradation, in the form of overgrazing, forest clearing or dense invasion by alien plants, on average reduces species populations by 40–60%. Cultivation and urbanization have a higher impact per unit land area than degradation, but a much smaller fraction of southern Africa is at risk.

Scalability and sensitivity of the index

A central feature of the BII is that it can be compared directly within and across scales. This was tested in practice by applying the BII to the three levels of environmental decision-making in South Africa: national ($1.2 \times 10^6 \text{ km}^2$), provincial (average area $1.35 \times 10^5 \text{ km}^2$) and local government (average area $4.6 \times 10^4 \text{ km}^2$), using biome-level species richness data compiled for South Africa³¹ in place of the WWF ecoregion richness data²⁷. The BII was found to deliver intuitively meaningful results at all scales.

Sensitivity to the resolution of species richness data was examined by comparing the values of BII for mammals obtained using individual mammal species distribution maps³², to that obtained using aggregated biome-level mammal richness data³¹. For South Africa overall, the BII for mammals was 67.5% using the biome-level data and 66.1% using the individual species data. At the level of individual biomes, the largest absolute difference between the two methods was 0.6%; the largest difference at both the provincial and municipal levels was 3.0%. No correlation was found between the calculated difference and the size of the units of aggregation.

Sensitivity to the type of species richness data was further explored by calculating 'functional biodiversity intactness' (as opposed to compositional biodiversity intactness⁷), in which the aggregated taxon-level I_{ijk} estimates and R_{ij} were derived by

weighting by functional group, instead of by species richness. Conceptually, this assigns greater weight to heavily affected, but relatively species-poor groups, such as megaherbivores, and negates the overwhelming effect of plants on the BII. The functional intactness (FII) of South Africa is 81.0%, and therefore slightly higher than the species-weighted BII (79.6%). Mammals showed by far the largest change (−9.7%) in comparison to the species-based estimate, reflecting the high impact of human activities on a few relatively species-poor functional groups. All other taxa showed small changes. These results suggest that in regions lacking species-level richness data, FII could serve as a first approximation for BII: I_{ijk} would be estimated on the basis of a sample of species per functional group, and R_{ij} would be a count of the number of functional groups present in a particular ecosystem type.

Sensitivity of the BII to the source of land use information was explored in two ways. Increasing or decreasing the total area of any single land use category by 5% of its current area resulted in an absolute change of less than 0.3% in the value of BII. The maximum compounded change across all categories was 0.7%. Second, the BII results based on the merged regional land use map (Table 1) were compared with those derived from the South African national land cover³³ and protected areas databases³⁴. The BII score derived from the national-level data is 81.2%, compared with a value of 79.6% using the regional land use map.

In both the case of species richness (R_{ij}) and land use (A_{jk}), the use of reasonable alternative data sources therefore resulted in a difference in the estimated BII (1.4% absolute for species richness and 1.6% for land use) of about twice the magnitude of the estimated decadal change in BII (0.8%), but five times smaller than the uncertainty associated with the impact factors I_{ijk} (6.4%). Thus, for purposes of documenting change, a single type of information source for species richness and land cover should be used for the beginning and end of the period, and the impact factors should be held constant.

Discussion

The BII is an indicator of the overall state of biodiversity in a given area, synthesizing land use, ecosystem extent, species richness and population abundance data. It is sensitive to the drivers and changes in the populations of species that typify the process of biodiversity loss, and robust to typical variations in data quality.

It is clear that a single index of biodiversity is not sufficient for all purposes. The BII is not intended to highlight individual species that are under threat, and should be used together with indicators such as the IUCN red list of threatened species. Conceptually, the BII is very similar to the natural capital index (NCI), which has been implemented in the Netherlands³⁵. However, the method for estimating BII does not require actual population data, and it can therefore complement the NCI in data-sparse regions.

The principal disadvantage of the BII is that it may be insensitive to slow acting, diffuse impacts on biodiversity, for instance the long-term effects of habitat fragmentation, climate change or pollution. However, an indicator does not have to be flawless to be useful. The gross domestic product is a good example: it is a universal standard

Table 2 BII (%) for southern Africa, per biome and taxonomic group

	Area (km ²)	Plants	Mammals	Birds	Reptiles	Amphibia	All taxa
Richness*		23,420	258	694	363	111	24,846
Forest	176,893	75.5	74.9	92.0	85.7	84.8	78.0
Savanna	2,329,550	85.5	73.2	95.5	88.9	95.9	87.0
Grassland	408,874	72.5	55.1	90.0	75.6	81.1	74.1
Shrubland	750,217	86.0	72.2	105.6	93.4	126.5	88.6
Fynbos	78,533	75.5	78.1	91.0	76.5	79.4	76.4
Wetland†	95,166	90.7	83.3	94.3	91.7	94.6	91.3
All biomes	3,839,233	82.4	71.3	96.0	88.1	95.1	84.4

* Or number of species. Data are for South Africa³¹, presented for indicative purposes only.

† Refers only to large wetlands, such as the Okavango Delta.

Table 3 BII calculated per country for southern Africa

Country	BII (%)	95% CI*
Botswana	88.5	±7.2
Lesotho	69.0	±7.1
Mozambique	89.5	±8.4
Namibia	91.4	±7.1
Swaziland	72.1	±7.1
South Africa	79.9	±6.5
Zimbabwe	76.2	±7.7
Region	84.4	±7.3

* The confidence interval reflects the uncertainty in the expert estimates, I_{ijk} .

for comparison, despite its many, well-known problems. Environmental policy could benefit from a 'macro-ecological' indicator of a similar level of generality. □

Received 17 November; accepted 15 December 2004; doi:10.1038/nature03289.

1. WCMC *Global Biodiversity: Earth's Living Resources in the 21st Century* (World Conservation Press, Cambridge, UK, 2000).
2. Anderson, J. A. (ed.) *Towards Gondwana Alive: Promoting Biodiversity and Stemming the Sixth Extinction* (Gondwana Alive Society, Pretoria, South Africa, 2001).
3. UN *World Summit on Sustainable Development: Johannesburg Plan of Implementation* (United Nations, New York, 2002).
4. The Royal Society *Measuring Biodiversity for Conservation* (The Royal Society, London, 2003).
5. Reid, W. V., McNeely, J. A., Tunstall, D. B., Bryant, D. A. & Winograd, M. *Biodiversity Indicators for Policy Makers* (World Resources Institute, Washington DC, 1993).
6. CBD *Proposed Biodiversity Indicators Relevant to the 2010 Target* (Convention on Biological Diversity, Montreal, 2003).
7. Noss, R. F. Indicators for monitoring biodiversity: A hierarchical approach. *Conserv. Biol.* **4**, 355–364 (1990).
8. CBD *Monitoring and Indicators: Designing National-Level Monitoring Programmes and Indicators* (Convention on Biological Diversity, Montreal, 2003).
9. Magurran, A. E. *Measuring Biological Diversity* (Blackwell, Oxford, 2004).
10. Hilton-Taylor, C. (ed.) *2000 IUCN Red List of Threatened Species* (IUCN, Gland, Switzerland and Cambridge, UK, 2000).
11. IUCN & UNEP *World Database on Protected Areas* (IUCN & UNEP, Cambridge, UK, 2003).
12. Loh, J. (ed.) *Living Planet Report 2002* (WWF International, Gland, Switzerland, 2002).
13. Sanderson, E. W. *et al.* The human footprint and the last of the wild. *Bioscience* **52**, 891–904 (2002).
14. Jenkins, M. Prospects for biodiversity. *Science* **302**, 1175–1177 (2003).
15. Houghton, J. T. *et al.* (eds) *Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories* (UK Meteorological Office, Bracknell, 1997).
16. Armstrong, A. J. & van Hensbergen, H. J. Comparison of avifaunas in *Pinus radiata* habitats and indigenous riparian habitat at Jonkershoek, Stellenbosch. *S. Afr. J. Wildl. Res.* **24**, 48–55 (1994).
17. Dean, W. R. J., Anderson, M. D., Milton, S. J. & Anderson, T. A. Avian assemblages in native *Acacia* and alien *Prosopis* drainage line woodland in the Kalahari, South Africa. *J. Arid Environ.* **51**, 1–19 (2002).
18. Fabricius, C., Burger, M. & Hockey, P. A. R. Comparing biodiversity between protected areas and adjacent rangeland in xeric succulent thicket, South Africa: arthropods and reptiles. *J. Appl. Ecol.* **40**, 392–403 (2003).
19. Herremans, M. Conservation status of birds in Botswana in relation to land use. *Biol. Conserv.* **86**, 139–160 (1998).
20. Johnson, R., Ferguson, J. W. H., van Jaarsveld, A. S., Bronner, G. N. & Chimimba, C. T. Delayed responses of small-mammal assemblages subject to afforestation-induced grassland fragmentation. *J. Mamm.* **83**, 290–300 (2002).
21. Joubert, D. F. & Ryan, P. G. Differences in mammal and bird assemblages between commercial and communal rangelands in the Succulent Karoo, South Africa. *J. Arid Environ.* **43**, 287–299 (1999).
22. Meik, J. M., Jeo, R. M., Mendelson, J. R. III & Jenks, K. E. Effects of bush encroachment on an assemblage of diurnal lizard species in central Namibia. *Biol. Conserv.* **106**, 29–36 (2002).
23. Parsons, D. A. B., Shackleton, C. M. & Scholes, R. J. Changes in herbaceous layer condition under contrasting land use systems in the semi-arid lowveld, South Africa. *J. Arid Environ.* **37**, 319–329 (1997).
24. Richardson, D. M. & van Wilgen, B. W. Effects of thirty-five years of afforestation with *Pinus radiata* on the composition of mesic mountain fynbos near Stellenbosch. *S. Afr. J. Bot.* **52**, 309–315 (1986).
25. Shackleton, C. M. Comparison of plant diversity in protected and communal lands in the Bushbuckridge lowveld savanna, South Africa. *Biol. Conserv.* **94**, 273–285 (2000).
26. Smart, R., Whiting, M. J. & Twine, W. Lizards and landscapes: Integrating field surveys and interviews to assess the impact of human disturbance on lizard assemblages and selected reptiles in a savanna in South Africa. *Biol. Conserv.* **122**, 23–31 (2005).
27. WWF-US *Species Database for African Ecoregions* (WWF Conservation Science Programme, Washington DC, 2003).
28. Olson, D. M. *et al.* Terrestrial ecoregions of the world: A new map of life on earth. *Bioscience* **51**, 933–938 (2001).
29. FAO *FAOSTAT Agricultural Database* (<http://apps.fao.org>) (2004).
30. FAO *Global Forest Resources Assessment 2000* (Food and Agriculture Organisation of the United Nations, Rome, 2001).
31. Le Roux, J. (ed.) *The Biodiversity of South Africa 2002: Indicators, Trends and Human Impacts* (Struik, Cape Town, 2002).
32. Friedmann, Y. & Daly, D. (eds) *Red Data Book of the Mammals of South Africa: a Conservation Assessment* (IUCN Conservation Breeding Specialist Group & Endangered Wildlife Trust, Johannesburg, South Africa, 2004).
33. CSIR *National Land Cover of South Africa* (Council for Scientific and Industrial Research, Division of Water, Environment and Forestry Technology, Pretoria, South Africa, 1996).
34. CSIR *Protected Areas of South Africa* (Department of Environmental Affairs and Tourism, Pretoria, South Africa, 2003).
35. Ten Brink, B. J. E. *et al.* *Technisch ontwerp Natuurwaarde 1.0 en toepassing in Natuurverkenning 2 [Technical design Natural Capital Index framework and implementation for the Nature Outlook 2]* (RIVM, Bilthoven, The Netherlands, 2002).
36. CSIR *SADC Landcover Dataset* (Council for Scientific and Industrial Research, Division of Water, Environment and Forestry Technology, Pretoria, South Africa, 2002).
37. EC *Global Land Cover 2000 (GLC2000) Database* (European Commission Joint Research Centre, Ispra, Italy, 2003).
38. CIESIN IFPRI & CIAT *Global Rural-Urban Mapping Project (GRUMP): Urban Extents* (CIESIN, Columbia Univ., Palisades, 2004).

Acknowledgements We thank the following experts whom we interviewed: G. Alexander, G. Bredenkamp, D. Biggs, B. Branch, V. Carruthers, A. Channing, C. Chimimba, J. du Toit, W. Haacke, J. Harrison, M. Keith, L. Minter, M. Rutherford, W. Tarboton and M. Whiting. R. Dean, C. Geldenhuys, B. Reyers, H. Simons, B. van Wilgen and M. Whiting are thanked for providing links to literature studies for validation of the expert estimates. We thank N. Burgess, J. D'Amico and the Conservation Science Programme at WWF, and G. Kier and the Biomap Working Group at the University of Bonn, Germany, for species richness data. J. Harrison is thanked for data used to derive the richness of each taxon functional type per biome. Thank you to D. van Zyl and T. Newby at the Institute for Soil, Climate and Water in Pretoria for the maximum annual normalized difference vegetation index (NDVI) data that were used to map degradation. B. Reyers at the University of Stellenbosch is thanked for the summation of the mammal distribution grids, which were provided by M. Keith. R.J.S. conceived the original idea for the index. It was then jointly developed further, and the paper was co-written by the authors. R.B. did the BII calculations, GIS application and literature validation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.J.S. (bscholes@csir.co.za).

A powerful bursting radio source towards the Galactic Centre

Scott D. Hyman¹, T. Joseph W. Lazio², Namir E. Kassim², Paul S. Ray³, Craig B. Markwardt⁴ & Farhad Yusef-Zadeh⁵

¹Department of Physics and Engineering, Sweet Briar College, Sweet Briar, Virginia 24595, USA

²Naval Research Laboratory, Code 7213, Washington, DC 20375-5320, USA

³E. O. Hulburt Center for Space Research, Naval Research Laboratory, Washington, DC 20375, USA

⁴Laboratory for High Energy Astrophysics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

⁵Northwestern University, Department of Physics and Astronomy, Evanston, Illinois 60208, USA

Transient astronomical sources are typically powered by compact objects and usually signify highly explosive or dynamic events¹. Although high-time-resolution observations are often possible in radio astronomy², they are usually limited to quite narrow fields of view. The dynamic radio sky is therefore poorly sampled, in contrast to the situation in the X-ray and γ -ray bands in which wide-field instruments routinely detect transient sources³. Here we report a transient radio source, GCRT J1745–3009, which was detected during a moderately wide-field monitoring programme of the Galactic Centre region^{4,5} at 0.33 GHz. The characteristics of its bursts are unlike those known for any other class of radio transient. If located in or near the Galactic Centre, its brightness temperature ($\sim 10^{16}$ K) and the implied energy density within GCRT J1745–3009 vastly exceed those observed in most other classes of radio astronomical sources⁶, and are consistent with coherent emission processes⁷ that are rarely observed. We conclude that it represents a hitherto unknown class of transient radio sources, the first of possibly many new classes that may be discovered by emerging wide-field radio telescopes⁸.

GCRT J1745–3009 is located approximately 1.25° south of the Galactic Centre and is notable for a series of ~ 1 -Jy bursts, each with a duration of ~ 10 min, and occurring at apparently regular intervals of 1.27 h. The variability of GCRT J1745–3009 is shown in the light curve of Fig. 1, and the average burst light curve is shown in Fig. 2. The light curves appear to be similar in shape, although the missing data during the first, second and third bursts hinders a comprehensive comparison. GCRT J1745–3009 is located near the supernova remnant, SNR 359.1–00.5 (ref. 9), and other prominent sources^{10,11}, as shown in Fig. 3. GCRT J1745–3009 was detected in 2002 using the Very Large Array radio telescope operating at a frequency of 0.33 GHz.

GCRT J1745–3009 is not detected between bursts with a 5σ upper limit of 75 mJy, determined by imaging the entire observation with the bursts removed. We also do not detect the source in 0.33 GHz, ~ 1 -h Galactic Centre monitoring observations made earlier in 2002 and afterwards in 2003; the 5σ upper limit for detection in a bursting state is ~ 250 mJy with 5-min integrations, and in a quiescent state is ~ 50 mJy. Images made from three 6-h observations in 1996 and 1998 have similar upper limits, and the combination of these images¹² has a 15-mJy upper limit for quiescent emission.

The magnitude of errors in radio astronomical images typically increases with distance from the centre of the image. GCRT J1745–3009 is located only $14'$ from the image centre compared to the $\sim 3^\circ$ field of view, and therefore, together with its detection at multiple frequencies around 0.33 GHz and in both circular polarizations, we consider the evidence that the source is real to be very strong. The bursts show no significant frequency dependence and no molecular-line masers are known to emit near 0.33 GHz; the lack of a frequency

dependence therefore suggests that GCRT J1745–3009 is not a maser.

GCRT J1745–3009 is unresolved in our observation. If we constrain its size to be less than $c\tau$, with c the speed of light in vacuum and $\tau \approx 2$ min taken to be the decay time of the ~ 1 -Jy bursts, then the energy density within the source as measured by the brightness temperature is $\sim 10^{12}$ K $(D/70 \text{ pc})^2$, where D is the distance to the source. If the transient source is at the Galactic Centre, ~ 8.5 kpc distant, its brightness temperature far exceeds 10^{12} K, the upper limit for incoherent synchrotron radiation⁶ produced by relativistic electrons gyrating in a magnetic field, and therefore its emission is probably coherent.

In principle, GCRT J1745–3009 could be located < 70 pc from us, in which case it could be either a coherent or an incoherent emitter. Known and hypothesized classes of 'local' ($D < 70$ pc) sources that show flare activity include dwarf M-type (dMe) stars, brown dwarfs and extrasolar planets. dMe flare stars emit coherent bursts produced through electron cyclotron maser emission. The bursts from flare stars show some similarities to the light curve of GCRT J1745–3009 but, in contrast, they are detected in only one circular polarization at low frequencies (for example, at 0.43 GHz for AD Leo and YZ Canis Minoris)¹³. Bursts of such highly circularly polarized radio emission are also predicted, by analogy to the giant planets in our Solar System, from extrasolar giant planets^{14,15}. However, no detections have been made in searches for such emission from known extrasolar planets at 0.33 GHz and 1.5 GHz, at sensitivity limits comparable to or better than what we report here¹⁶. We conclude that GCRT J1745–3009 is not likely to be a dMe flare star or an extrasolar planet.

Brown dwarfs also emit flares, apparently as a result of processes involving high magnetic fields¹⁷. Four infrared sources detected in

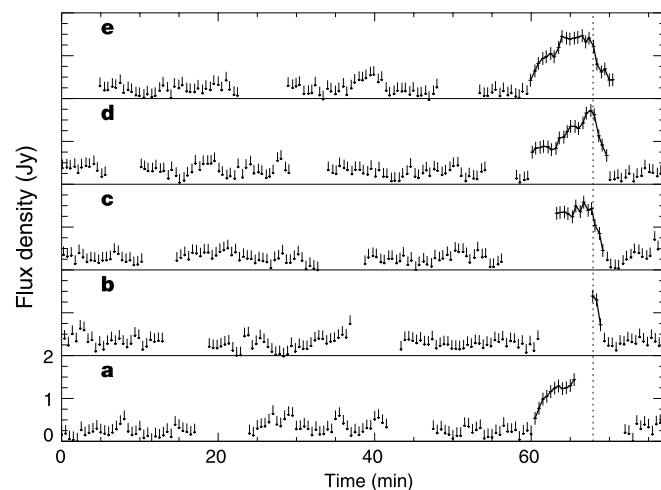


Figure 1 The five detected bursts from the radio transient source, GCRT J1745–3009. The observation is continuous, with the time axis folded at multiples of 77.13 min. **a**, The first interval of the observation, beginning at 20 h 50 min 00 s on 30 September 2002 (TAI). **b**, The second interval. **c**, The third interval. **d**, The fourth interval. **e**, The fifth interval. The points connected by the heavy line are the detections in 30-s samples with typical error bars of 0.15 Jy shown. The arrows are 3σ upper limits for nondetections between bursts; no evidence of quiescent emission is found. Fluctuations of nearby sources are consistent with the noise level. The dashed vertical line is positioned at the fitted peak (see Fig. 2) of the fourth burst as a reference. Note that several gaps in the data, including during the first three bursts, are due to radio-frequency interference or when the phase calibrator was imaged. No anomalous behaviour is seen for the calibrator. The 0.33-GHz, 7-h observation was obtained with the CnB-configuration (~ 40 arcsec resolution) of the Very Large Array. The bandpass consists of 31 97-kHz-wide channels for each of two intermediate frequencies (321.56 and 327.50 MHz). Both circular polarizations were imaged, but linear polarization measurements are not available for this observation. No circular polarization was detected (15%, 5σ upper limit).

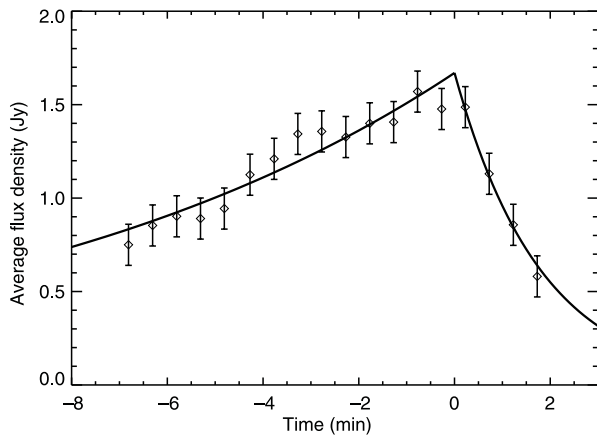


Figure 2 Average light curve of GCRT J1745–3009 derived from the third, fourth and fifth bursts. The peak times of the bursts, t_0 , were determined for all but the second, most incompletely sampled burst, from exponential fits to their rise ($Ae^{(t-t_0)/\tau_1}$) and decay ($Ae^{-(t-t_0)/\tau_2}$) and have been subtracted from the light curve of each to construct the average. (Although the first burst was included in the fit, it is also poorly sampled and is therefore not included in the average light curve.) The peak of the average light curve is arbitrarily placed at 0 min. Typical error bars (± 0.11 Jy) are shown. The amplitude, A , and rise and decay time constants, τ_1 and τ_2 , were constrained to be identical in the fits, yielding $A = 1.67 \pm 0.05$ Jy, $\tau_1 = 9.9 \pm 0.7$ min, and $\tau_2 = 1.9 \pm 0.5$ min. The difference between the peak times derived for the first and third bursts is $2 \times 77.3 \pm 0.2$ min, the third and fourth 77.6 ± 0.3 min, and the fourth and fifth 76.5 ± 0.3 min. A separate fit (solid curve) was made to the average light curve and yields parameters consistent with those listed above.

the Two-Micron All-Sky Survey¹⁸ that lie within the $\sim 10''$ uncertainty of the transient's radio position could possibly be brown dwarfs. However, their distances are unknown and their spectral colours, to the extent measured, are inconsistent with those of brown dwarfs¹⁹.

Few radio flares from brown dwarfs have been detected^{20–22}, and

none at low frequencies¹⁶. Like the bursts from GCRT J1745–3009, the observed flares have timescales of minutes, and, for two of the brown dwarfs, LP944–20 and 2MASS 0036+18, they also have faster decay than rise times. In contrast to GCRT J1745–3009, however, no regular flaring pattern is evident for any of these sources, and they were observed only at high frequencies, because emission is predicted to be self-absorbed at lower frequencies²³. The extent to which flares from brown dwarfs are self-absorbed at low frequencies is somewhat uncertain, though, because flares from only two objects (LP944–20 and DENIS 1048–3956) have been observed at multiple frequencies simultaneously, and these frequencies were much higher (4.8 and 8.5 GHz) than 0.33 GHz, at which we detect GCRT J1745–3009. Furthermore, in contrast to the low-frequency bursts from GCRT J1745–3009, the observed flares from brown dwarfs at high frequencies are significantly polarized (30–70%), but the degree of polarization may be significantly reduced at low frequencies where self-absorption can become large²⁴.

We conclude, therefore, that we cannot rule out that GCRT J1745–3009 is a flaring brown dwarf. However, its detection at a low frequency, its regular bursting pattern, and significant detection in both circular polarizations, are all novel features in comparison to known characteristics of brown dwarfs. Hence, there is no compelling evidence for such an identification.

Although GCRT J1745–3009 is conceivably a 'local' radio source belonging to one of the classes of objects considered above, it is much more likely that it is located significantly farther from us. Assuming even a uniform distribution of transients, and not the vastly increasing spatial density of all astronomical objects toward the Galactic Centre, the relative spatial volume covered by our wide-field observations results in an extremely small probability (6×10^{-7}) that GCRT J1745–3009 is located within 70 pc.

Many transient radio sources are also detectable at X-ray and γ -ray wavelengths. We have analysed a serendipitous pointed observation about $32'$ from GCRT J1745–3009 in the Rossi X-Ray Timing Explorer (RXTE) archive that runs between the third and fourth burst, although it overlaps neither. No variable X-ray

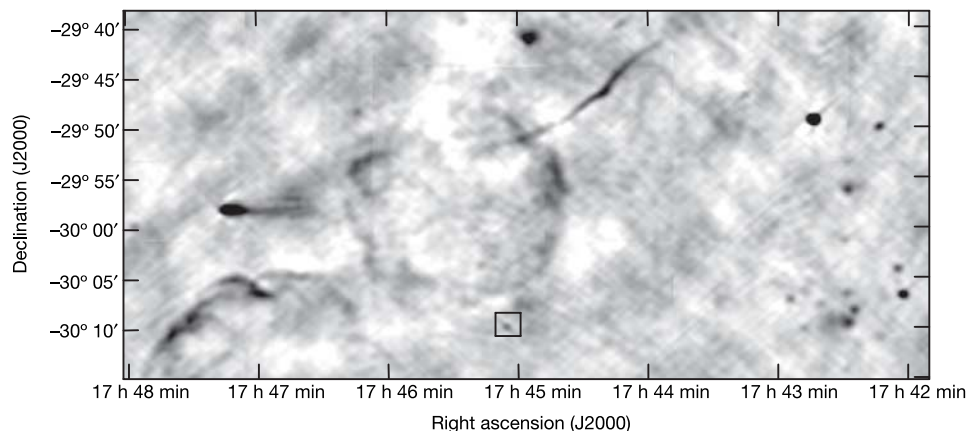


Figure 3 0.33 GHz radio image of the transient source, GCRT J1745–3009, and the surrounding region $\sim 1^\circ$ south of the Galactic Centre. GCRT J1745–3009 is located at (J2000) right ascension $17\text{ h }45\text{ min }5 \pm 0.8\text{ s}$, declination $-30^\circ 09' 52'' \pm 10''$, indicated by the small box below the $\sim 20'$ -diameter shell of the supernova remnant, SNR 359.1–00.5. Other sources in the image include the sources to the west which are part of Sagittarius E, the linear feature, 'The Snake', to the north, and 'The Mouse' to the northeast of GCRT J1745–3009. The sensitivity and resolution of the image are 15 mJy beam^{-1} and $48'' \times 39''$, respectively. We searched the entire 3° field of view for other intra-observation transients and variable sources and found none. Note that GCRT J1745–3009 appears as only a 100-mJy source here because it is averaged over the five short ~ 1 -Jy bursts and a total of ~ 6 -h of nondetections between bursts. Nondetections of quiescent and bursting emission at 0.33 GHz for other epochs are described in the text.

We also do not detect quiescent emission in a 1.4-GHz observation from January 2003, with a detection threshold of 35 mJy. However, if GCRT J1745–3009 has a steep spectrum, as found for other radio transients (for example, $\alpha = -1.2$, $S_\nu \propto \nu^\alpha$ for the Galactic Center Transient³⁰) and coherent emitters such as radio pulsars ($\alpha \approx -1.7$), it easily could have decayed significantly below the January 2003 detection threshold. Also, although no bursts were detected within the 1.4-GHz observations, these data were comprised of seven 3-min 'snapshots' taken every hour; given its 1.27-h recurrence interval, GCRT J1745–3009 in a bursting state could easily have been missed. Similarly, most of the 0.33-GHz nondetection observations from 2002 and 2003 were only about an hour or less in duration, and therefore it is possible that GCRT J1745–3009 was and still is detectable in a bursting state. North is at the top of the image.

(2–10 keV) emission was seen and we place a conservative 25 mCrab upper limit on the flux of any interburst X-ray counterpart emission ($1 \text{ mCrab} \approx 2 \times 10^{-11} \text{ erg cm}^{-2} \text{ s}^{-1}$). One of us (C.B.M.) performs regular scanning observations of the Galactic bulge with RXTE. The closest of these were on 25 September and 2 October 2002. In both of these observations, 3σ upper limits on the X-ray flux from the source are 6 mCrab. A search of all of the Galactic bulge scans starting in February 1999 finds one scan, on 3 July 2003, in which a 15-mCrab (6.5σ) outburst was detected. However, possible confusion with other X-ray sources in the field of view ($\sim 30'$) prevents a conclusive identification. Similarly, the γ -ray source 3EG J1744–301 was detected near the position of GCRT J1745–3009 during the 1990s by the Energetic Gamma Ray Experiment Telescope, but the positional error on that source is also large ($\sim 20'$) and the source is in a highly confused region with numerous diffuse and discrete sources of γ -ray emission²⁵.

We next consider the possibility that the bursts from GCRT J1745–3009 could be relativistically beamed toward us, as is the case for microquasars, which are accreting black holes in binary systems that occasionally power radio-bright relativistic jets (for example GRS 1915+105, whose apparent superluminal motion has a Lorentz factor of $\gamma \approx 5$)^{26,27}. It is conceivable that relativistic beaming is responsible for the high calculated brightness temperature, but the light curve of GCRT J1745–3009 does not resemble that for known microquasars or other sources of jet emission, most of which exhibit a fast rise and a slower decay and much longer timescales³. In addition, the apparent lack of a bright X-ray counterpart of the bursts argues strongly against accretion as the power source for the bursts.

Next, we consider radio pulsar origins for the source. A 77-min rotation period radio pulsar is excluded because the rotational energy loss rate from such a pulsar is insufficient to power the radio emission unless the magnetic field is extreme ($>10^{18} \text{ G}$) or the distance is unreasonably small ($<0.5 \text{ pc}$). Another conceivable option is that 77 min is an orbital period and the outbursts are flux variations as a function of orbital phase similar to the pulsar PSR J0737–3039B (ref. 28). This scenario does not explain the transient behaviour from the source or the lack of interburst emission, and favours a distance of the order of 1 kpc or less.

One other class of sources to consider are magnetars, neutron stars with immense (10^{14} – 10^{15} G) magnetic fields whose radiation is powered by field decay²⁹. Coherent emission from the magnetosphere of a magnetar addresses the energy budget difficulties seen in the pulsar models, but all known magnetars have much shorter spin periods ($\sim 10 \text{ s}$) than the 77-min period observed for GCRT J1745–3009. An investigation of whether a new, long-period type of magnetar could produce the observed emission timescales and transient behaviour is under way (K. S. Wood, P.S.R., S.D.H., T.J.W.L. & N.E.K., manuscript in preparation). □

Received 26 November 2004; accepted 24 January 2005; doi:10.1038/nature03400.

1. Cordes, J. M., Lazio, T. J. W. & McLaughlin, M. A. The dynamic radio sky. Preprint at (<http://arxiv.org/abs/astro-ph/0410045>) (2004).
2. Hankins, T. H., Kern, J. S., Weatherall, J. C. & Eilek, J. A. Nanosecond radio bursts from strong plasma turbulence in the Crab pulsar. *Nature* **422**, 141–143 (2003).
3. Bradt, H., Levine, A., Remillard, R. & Smith, D. A. in *AIP Conf. Proc. on X-ray Astronomy: Stellar Endpoints, AGN, and the Diffuse X-ray Background* (eds White, N. E., Malaguti, G. & Palumbo, G. G. C.) **599**, 35–52 (2001).
4. Hyman, S. D., Lazio, T. J. W., Kassim, N. E. & Bartleson, A. L. Low-frequency radio transients in the galactic center. *Astron. J.* **123**, 1497–1501 (2002).
5. Hyman, S. D., Lazio, T. J. W., Kassim, N. E., Nord, M. E. & Neureuther, J. L. A search for radio transients at 0.33 GHz in the G.C. *Astron. Nachr.* **324**(S1), 79–83 (2003).
6. Readhead, A. C. S. Equipartition brightness temperature and the inverse Compton catastrophe. *Astrophys. J.* **426**, 51–59 (1994).
7. Melrose, D. B. Coherent emission in AGN: a critique. *Publ. Astron. Soc. Aust.* **19**, 34–38 (2002).
8. The long wavelength array. (<http://lwa.nrl.navy.mil>) (Naval Research Laboratory, 2004).
9. Reich, W. & Fürst, E. G357.7+0.3 and G359.1–0.5—two shell-type supernova remnants in the galactic centre region. *Astron. Astrophys. Suppl. Ser.* **57**, 165–167 (1984).
10. Yusef-Zadeh, F. & Bally, J. A non-thermal axially symmetric radio wake towards the galactic center. *Nature* **330**, 455–458 (1987).
11. Gray, A. D., Cram, L. E., Ekers, R. D. & Goss, W. M. A filamentary radio source near the galactic centre. *Nature* **353**, 237–239 (1991).

12. Nord, M. E. *et al.* High-resolution, wide-field imaging of the galactic center region at 330 MHz. *Astron. J.* **128**, 1646–1670 (2004).
13. Bastian, T. S., Bookbinder, J., Dulk, G. A. & Davis, M. Dynamic spectra of radio bursts from flare stars. *Astrophys. J.* **353**, 265–273 (1990).
14. Farrell, W. M., Desch, M. D. & Zarka, P. On the possibility of coherent cyclotron emission from extrasolar planets. *J. Geophys. Res.* **104**, 14025–14032 (1999).
15. Lazio, T. J. W. *et al.* The radiometric Bode's Law and extrasolar planets. *Astrophys. J.* **612**, 511–518 (2004).
16. Bastian, T. S., Dulk, G. A. & Leblanc, Y. A search for radio emission from extrasolar planets. *Astrophys. J.* **545**, 1058–1063 (2000).
17. Guedel, M. Stellar radio astronomy: probing stellar atmospheres from protostars to giant. *Annu. Rev. Astron. Astrophys.* **40**, 217–261 (2002).
18. Cutri, R. M. *et al.* The 2MASS All-sky Catalog of Point Sources (University of Massachusetts and Infrared Processing and Analysis Center, IPAC/California Institute of Technology, 2003).
19. Burgasser, A. J. *et al.* The spectra of T dwarfs. I. Near-infrared data and spectral classification. *Astrophys. J.* **564**, 421–451 (2002).
20. Berger, E. *et al.* Discovery of radio emission from the brown dwarf LP944–20. *Nature* **410**, 338–340 (2001).
21. Berger, E. Flaring up all over-radio activity in rapidly rotating late M and L dwarfs. *Astrophys. J.* **572**, 503–513 (2002).
22. Putman, M. E. & Burgasser, A. J. Radio emission from late-type dwarfs: quiescent emission and a spectacular radio flare from the M9 DENIS 1048–3956. *Bull. Am. Astron. Soc.* **35**, Abstr. 43.07 (2003); (<http://www.aas.org/publications/baas/v35n5/aas203/1159.htm>).
23. Dulk, G. A. & Marsh, K. A. Simplified expressions for the gyrosynchrotron radiation from mildly relativistic, nonthermal and thermal electrons. *Astrophys. J.* **259**, 350–358 (1982).
24. Bruggemann, G. & Magun, A. Temporal and spectral characteristics of the circular polarization of solar microwave bursts. *Astron. Astrophys.* **239**, 347–355 (1990).
25. Mayer-Hasselwander, H. A. *et al.* High-energy gamma-ray emission from the Galactic Center. *Astron. Astrophys.* **335**, 161–172 (1998).
26. Marscher, A. P. *et al.* Observational evidence for the accretion-disk origin for a radio jet in an active galaxy. *Nature* **417**, 625–627 (2002).
27. Fender, R. P. *et al.* Merlino observations of relativistic ejections from GRS 1915+105. *Mon. Not. R. Astron. Soc.* **304**, 365–376 (1999).
28. Lyne, A. G. *et al.* A double-pulsar system: a rare laboratory for relativistic gravity and plasma physics. *Science* **303**, 1153–1157 (2004).
29. Woods, P. M. & Thompson, C. Soft gamma repeaters and anomalous X-ray pulsars: magnetar candidates. In *Compact Stellar X-ray Sources* (eds Lewin, W. H. G. & van der Klis, M.) (in the press); preprint at (<http://arxiv.org/astro-ph/0406133>).
30. Zhao, J.-H. *et al.* A transient radio source near the center of the milky way galaxy. *Science* **255**, 1538–1543 (1992).

Acknowledgements S.D.H. thanks J. Neureuther and M. Lazarova for their assistance in the Galactic Centre transient monitoring programme. We thank D. Chakrabarty, G. Denn, C. Dermer, W. Erickson, R. Remillard, K. Weiler and K. Wood for discussions. The Very Large Array (VLA) is operated by the National Radio Astronomy Observatory (NRAO), which is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities. Basic research in radio astronomy is supported at Sweet Briar College by funding from the Jeffress Memorial Trust and Research Corporation. Basic research in radio and X-ray astronomy at the NRL is supported by the Office of Naval Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.D.H. (shyman@sbc.edu).

Plasma formation and temperature measurement during single-bubble cavitation

David J. Flannigan & Kenneth S. Suslick

Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Single-bubble sonoluminescence (SBSL^{1–5}) results from the extreme temperatures and pressures achieved during bubble compression; calculations have predicted^{6,7} the existence of a hot, optically opaque plasma core⁸ with consequent bremsstrahlung radiation^{9,10}. Recent controversial reports^{11,12} claim the observation of neutrons from deuterium–deuterium fusion during acoustic cavitation^{11,12}. However, there has been previously no strong experimental evidence for the existence of a

plasma during single- or multi-bubble sonoluminescence. SBSL typically produces featureless emission spectra¹³ that reveal little about the intra-cavity physical conditions or chemical processes. Here we report observations of atomic (Ar) emission and extensive molecular (SO) and ionic (O_2^+) progressions in SBSL spectra from concentrated aqueous H_2SO_4 solutions. Both the Ar and SO emission permit spectroscopic temperature determinations, as accomplished for multi-bubble sonoluminescence with other emitters^{14–16}. The emissive excited states observed from both Ar and O_2^+ are inconsistent with any thermal process. The Ar excited states involved are extremely high in energy (>13 eV) and cannot be thermally populated at the measured Ar emission temperatures (4,000–15,000 K); the ionization energy of O_2 is more than twice its bond dissociation energy, so O_2^+ likewise cannot be thermally produced. We therefore conclude that these emitting species must originate from collisions with high-energy electrons, ions or particles from a hot plasma core.

In order to gain insight into the physical conditions and chemical processes occurring during single-bubble cavitation, we have explored a variety of low-volatility liquids¹⁷, observing both stationary and moving SBSL. We have now discovered the generation of extremely intense SBSL from a moving single bubble in concentrated aqueous H_2SO_4 solutions, $H_2SO_4(aq.)$, which have very low vapour pressures¹⁸ and are essentially transparent down to a wavelength of 200 nm. Under optimal conditions, the SBSL intensity from 85 wt% $H_2SO_4(aq.)$ under Ar is increased 2,700 times compared to that from water under Ar, and the intensity under Xe is increased 1,500 times compared to water under Xe (Fig. 1). When Xe is dissolved in degassed 85% $H_2SO_4(aq.)$, the SBSL is, by more than two orders of magnitude, the most intense yet observed in any liquid. SBSL from H_2SO_4 solutions had been previously reported, but with only a few-fold increase in intensity compared to water¹⁹.

Bright SBSL spectra from concentrated $H_2SO_4(aq.)$ are similar to water SBSL spectra, consisting of a featureless continuum that increases towards the ultraviolet (UV). Although these spectra are much more intense, blackbody fits (which have been previously reported for SBSL in water²⁰) do not yield vastly different temperatures for the SBSL now observed from $H_2SO_4(aq.)$; all four

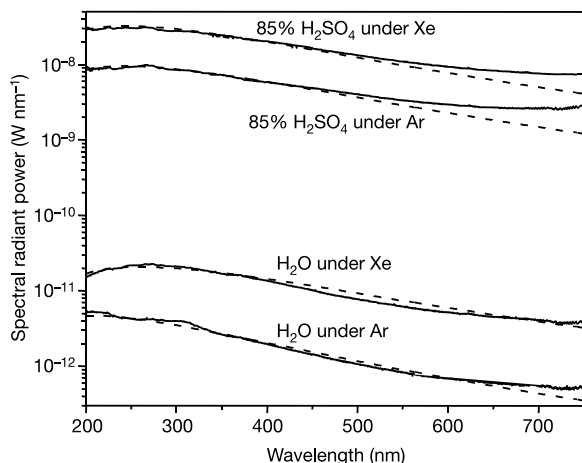


Figure 1 SBSL spectra from 85% $H_2SO_4(aq.)$ and pure water regassed with Xe and Ar (solid lines); apparent fits to blackbody spectra are given as dashed lines. Noble gas content of solutions, acoustic frequency and acoustic pressure were adjusted to obtain the brightest SBSL emission. The SBSL apparatus, spectral acquisition system and acoustic pressure measurements have been previously described¹⁷. All spectra were corrected for solution and resonator absorption, and for the response of the optical system (against NIST traceable standard lamps).

spectra in Fig. 1 have an apparent blackbody temperature of $\sim 12,500 \pm 1,500$ K. The very marked increase in SBSL intensity seen from concentrated $H_2SO_4(aq.)$ is due in part to the much lower concentration of polyatomic species inside the bubble relative to water at the same temperature: less of the energy of cavitation is consumed by endothermic bond dissociations²¹. The remaining bubble contents, comprised almost entirely of noble gas atoms mixed with a small amount of H_2SO_4 vapour, must therefore absorb a larger portion of the energy of collapse.

Importantly, the parameter space that supports SBSL in $H_2SO_4(aq.)$ is much larger than in water, with SBSL observable over an acoustic pressure range of 1.3 to >6 bar. (The acoustic pressures are estimated from hydrophone measurements at the centre of the acoustic field; movement of the bubble about the

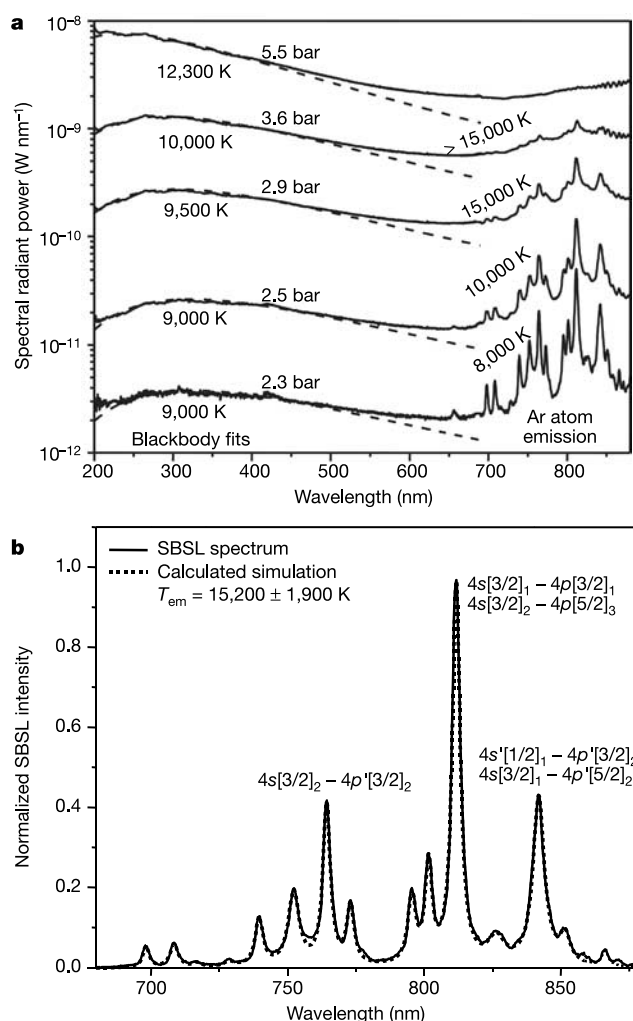


Figure 2 SBSL spectra from 85% $H_2SO_4(aq.)$. **a**, Solid lines are the observed SBSL emission spectra, dashed lines are calculated blackbody spectra. Applied acoustic pressures are shown above their corresponding plot. Temperatures of blackbody fits and of Ar atom emission (compare with **b**) are shown next to their corresponding plot. The strong neutral Ar atom emission lines in the red and near-infrared region of the spectra arise from the $4s-4p$ manifold. The $4s$ state is situated 11.5–11.8 eV above the ground state ($3p$), whereas the $4p$ state lies 13.1–13.5 eV above the ground state. **b**, Ar atom line emission (2.8 bar; solid line) compared to a calculated Ar atom emission spectrum at 15,200 K (dashed line). The underlying continuum has been subtracted; spectra are normalized to intensity at 811 nm. The most prominent lines are labelled with the corresponding transitions. The lines at 811 and 840 nm each contain contributions from two relatively strong transitions.

pressure antinode will decrease the effective acoustic pressure experienced by the bubble, so acoustic pressures given here are upper limits.) In SBSL from more weakly driven bubbles, for example, we observe very strong atomic emission from neutral Ar (Fig. 2). In this case, as the driving acoustic pressure is increased, the intensity of Ar emission lines decreases relative to the continuum emission, and the Ar lines broaden gradually into unresolved peaks at >5 bar. Observation of emission from excited states of atomic Ar allows us to calculate emission temperatures from the well-known energy levels, transition probabilities, statistical weights, and photon energies of Ar (ref. 22). Ar atom emission temperatures generated during SBSL from 85% $\text{H}_2\text{SO}_4(\text{aq.})$ were observed to increase with increasing acoustic pressure, and were found to be $8,000 \pm 1,000$ K at 2.3 bar, $10,700 \pm 1,300$ K at 2.5 bar, and $15,200 \pm 1,900$ K at 2.8 bar. Note that these experimentally determined temperatures are consistent with theoretically predicted SBSL temperatures^{7,23}. At higher acoustic pressures, accurate fits to Ar emission spectra become difficult owing to line broadening and the development of line asymmetries, possibly due to increased Ar ion concentrations. At the highest acoustic pressures reached before the bubble was forced from the centre of the resonator (>6 bar), calculation of SBSL temperature was impossible owing to the extreme broadening of all spectral features.

We note especially that the Ar excited states that are being populated are extremely high in energy (>13 eV)²⁴, which is too high to be populated thermally at the <1 eV effective emission temperature of Ar. This provides, to our knowledge, the first experimental evidence for excitation via high-energy particle collisions (for example, electron impact on Ar) from a hot plasma core. The possibility of an optically opaque, highly ionized plasma core during SBSL has been previously suggested by calculations⁸. Similar circumstances occur during intense shockwave heating in gases. For example, the limiting emission temperature of an intense shockwave in air is only $\sim 17,000$ K (ref. 25), which represents the opacity limit of shock heating of air, and shock-heated Ar plasmas typically have line emission temperatures below $\sim 20,000$ K (ref. 26).

Comparisons of Ar atom emission temperatures to blackbody fits of the UV continuum (Figs 1 and 2a) are problematic for several reasons. First, the origin of the UV continuum emission is unclear:

it may be partially blackbody radiation from the emitting shell surrounding an optically opaque core^{6,7}, and partially bremsstrahlung due to ionization^{9,10,27}. It could also be due to ion–electron recombination, which is likely to be present in a dense plasma^{9,10}. Furthermore, the continuum emission and the well-resolved Ar atom emission do not necessarily originate at the same time during bubble collapse or from the same spatial region within the bubble. A simple blackbody fit under these circumstances is unlikely to produce an accurate temperature measurement, and further comparison is therefore unwarranted.

According to current theoretical models of SBSL, compressional heating of the bubble should show some dependence on the thermal conductivity of the gas within the bubble^{23,28}. As shown in Fig. 3, we can use mixtures of Ar and Ne to systematically control the emission temperature from $>15,000$ K down to $\sim 1,500$ K. The effects of thermal conductivity may come both from direct thermal transport and from changes in the size of the plasma core^{8,23,25}. Thus it is clear that bubble collapse is only approximately adiabatic, and that even in SBSL, thermal conductivity of the dissolved gases is a critical experimental parameter.

We also observe extensive vibronic progressions arising from the $\text{B}^3\Sigma^- - \text{X}^3\Sigma^-$ system of sulphur monoxide (SO; ref. 29) in the SBSL spectra of concentrated $\text{H}_2\text{SO}_4(\text{aq.})$. Sulphur monoxide vibronic progressions are most prominent when the $\text{H}_2\text{SO}_4(\text{aq.})$ solution contains dissolved Ne (Fig. 4a). The bands are very weakly present in SBSL spectra of a very dimly luminescing Ar bubble, and they are totally absent from SBSL spectra of Kr and Xe bubbles, no matter how dim. If heat is rapidly conducted out of a Ne-filled bubble during collapse, dissociation of SO will occur to a lesser degree, resulting in an increase in SO emission intensity, as observed.

By comparing relative intensities of inter-progression SO emission bands, and also by observing how relative intensities of the bands change with increasing acoustic pressure, effective SO temperatures were determined to be $1,580 \pm 110$ K at 3.3 bar, $2,470 \pm 170$ K at 4.2 bar, and $3,480 \pm 240$ K at 5.1 bar, all regassed with Ne. At higher acoustic pressures, the emission temperatures continue to increase and become difficult to determine, owing to the blending of bands and the decrease in SO emission intensity (as dissociation occurs). Note that at the highest acoustic pressures applied (>6 bar), features of SO molecular emission were still apparent in the spectra. Emission temperatures from SO are of course limited by its dissociation, and so may represent either

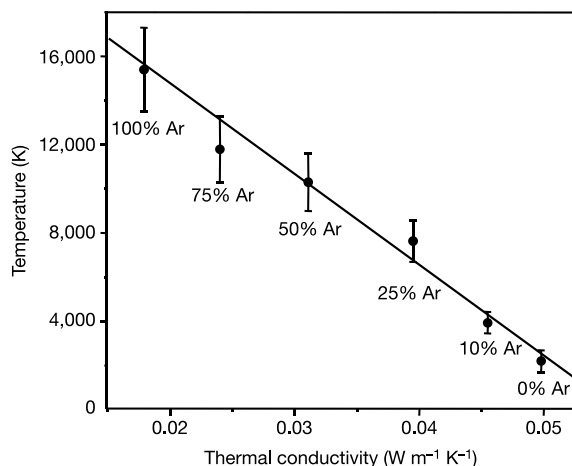


Figure 3 Emission temperatures of SBSL of 85% $\text{H}_2\text{SO}_4(\text{aq.})$ regassed with Ar/Ne mixtures (acoustic pressure 3 bar) are shown as a function of the thermal conductivity of the gas mixtures. Temperatures were calculated from Ar atom emission for all but the 100% Ne measurement, whose temperature was determined from SO emission (compare Fig. 4). Error bars are determined from the errors in the transition probabilities reported in ref. 20, and from an estimated 5% experimental error in the measured temperature.

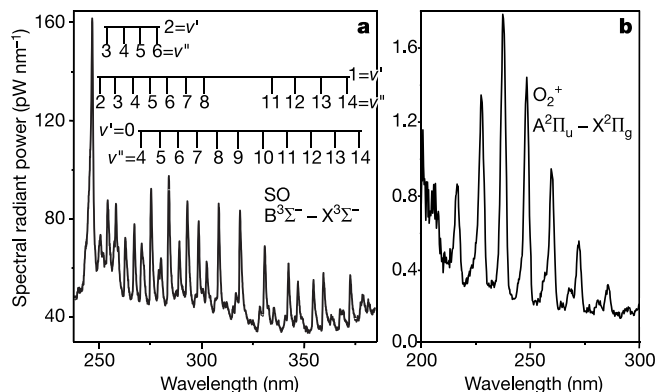


Figure 4 Vibronic progressions in SBSL spectra. **a**, Sulphur monoxide (SO) ($\text{B}^3\Sigma^- - \text{X}^3\Sigma^-$) emission from 85% $\text{H}_2\text{SO}_4(\text{aq.})$ regassed with Ne. The intense peak near 250 nm arises from the SO ultraviolet system ($\text{A}^3\Pi - \text{X}^3\Sigma^-$; ref. 29). Vibrational levels in the $\text{B}^3\Sigma^-$ state are denoted by v' while vibrational levels in the $\text{X}^3\Sigma^-$ state are denoted by v'' . **b**, Dioxygen cation (O_2^+) ($\text{A}^2\Pi_u - \text{X}^2\Pi_g$) emission³⁰ from 85% $\text{H}_2\text{SO}_4(\text{aq.})$ regassed with O_2/Xe .

emission from outer regions of the collapsing bubble or early (or late) times during bubble collapse.

Confirmation of the presence of a plasma in the collapsing bubble comes from the observation of O_2^+ emission³⁰ (Fig. 4b). The bond energies of O_2 and O_2^+ are 5.1 and 6.5 eV, respectively, while the ionization energy of O_2 is much higher at 12.1 eV. A total of over 18 eV of energy is necessary to form excited O_2^+ . The formation and excitation of O_2^+ therefore cannot occur thermally, and probably occurs via high-energy electron impact from the hot, opaque plasma core. To our knowledge, this is the first example of emission from any ion in any sonoluminescence spectrum under any conditions. □

Received 15 September 2004; accepted 17 January 2005; doi:10.1038/nature03361.

- Barber, B. P. & Putterman, S. J. Light scattering measurements of the repetitive supersonic implosion of a sonoluminescing bubble. *Phys. Rev. Lett.* **69**, 3839–3842 (1992).
- Gomph, B., Günther, R., Nick, G., Pecha, R. & Eisenmenger, W. Resolving sonoluminescence pulse width with time-correlated single photon counting. *Phys. Rev. Lett.* **79**, 1405–1408 (1997).
- Gaitan, D. F., Crum, L. A., Church, C. C. & Roy, R. A. Sonoluminescence and bubble dynamics for a single, stable, cavitation bubble. *J. Acoust. Soc. Am.* **91**, 3166–3183 (1992).
- Lohse, D., Brenner, M. P., Dupont, T. F., Hilgenfeldt, S. & Johnston, B. Sonoluminescing air bubbles rectify argon. *Phys. Rev. Lett.* **78**, 1359–1362 (1997).
- Brenner, M. P., Hilgenfeldt, S. & Lohse, D. Single-bubble sonoluminescence. *Rev. Mod. Phys.* **74**, 425–484 (2002).
- Moss, W. C., Clarke, D. B. & Young, D. A. Calculated pulse widths and spectra of a single sonoluminescing bubble. *Science* **276**, 1398–1401 (1997).
- Moss, W. C. *et al.* Computed optical emissions from a sonoluminescing bubble. *Phys. Rev. E* **59**, 2986–2992 (1999).
- Burnett, P. D. S. *et al.* Modeling a sonoluminescing bubble as a plasma. *J. Quant. Spectrosc. Radiat. Transfer* **71**, 215–223 (2001).
- Hilgenfeldt, S., Grossmann, S. & Lohse, D. A simple explanation of light emission in sonoluminescence. *Nature* **398**, 402–405 (1999).
- Yasui, K. Mechanism of single-bubble sonoluminescence. *Phys. Rev. E* **60**, 1754–1758 (1999).
- Taleyarkhan, R. P. *et al.* Evidence for nuclear emissions during acoustic cavitation. *Science* **295**, 1868–1873 (2002).
- Taleyarkhan, R. P. *et al.* Additional evidence of nuclear emissions during cavitation. *Phys. Rev. E* **69**, 036109 (2004).
- Hiller, R., Weninger, K., Putterman, S. J. & Barber, B. P. Effect of noble gas doping in single-bubble sonoluminescence. *Science* **266**, 248–250 (1994).
- McNamara, W. B. III, Didenko, Y. T. & Suslick, K. S. Sonoluminescence temperatures during multi-bubble cavitation. *Nature* **401**, 772–775 (1999).
- Flint, E. B. & Suslick, K. S. The temperature of cavitation. *Science* **253**, 1397–1399 (1991).
- Didenko, Y. T., McNamara, W. B. III & Suslick, K. S. Effect of noble gases on sonoluminescence temperatures during multi-bubble cavitation. *Phys. Rev. Lett.* **84**, 777–780 (2000).
- Didenko, Y. T., McNamara, W. B. III & Suslick, K. S. Molecular emission from single-bubble sonoluminescence. *Nature* **407**, 877–879 (2000).
- Greenewalt, C. H. Partial pressures of aqueous solutions of sulfuric acid. *J. Ind. Eng. Chem.* **17**, 522–523 (1925).
- Troia, A., Ripa, D. M. & Spagnolo, R. in *World Congress on Ultrasonics* (ed. Cassereau, D.) 1041–1044 (Société Française d'Acoustique, Paris, 2003).
- Vazquez, G., Camara, C., Putterman, S. & Weninger, K. Sonoluminescence: Nature's smallest blackbody. *Opt. Lett.* **26**, 575–577 (2001).
- Didenko, Y. T. & Suslick, K. S. The energy efficiency of formation of photons, radicals and ions during single-bubble cavitation. *Nature* **418**, 394–397 (2002).
- Wiese, W. L., Brault, J. W., Danzmann, K., Helbig, V. & Kock, M. Unified set of atomic transition probabilities for neutral argon. *Phys. Rev. A* **39**, 2461–2471 (1989).
- Toegel, R. & Lohse, D. Phase diagrams for sonoluminescing bubbles: A comparison between experiment and theory. *J. Chem. Phys.* **118**, 1863–1875 (2003).
- Cooper, R., Grieser, F., Sauer, M. C. Jr & Sangster, D. F. Formation and decay kinetics of the 2p levels of neon, argon, krypton, and xenon produced by electron-beam pulses. *J. Phys. Chem.* **81**, 2215–2220 (1977).
- Zel'dovich, Y. B. & Raizer, Y. P. *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Academic, New York, 1966).
- Tourin, R. H. *Spectroscopic Gas Temperature Measurement* (Elsevier, Amsterdam, 1966).
- Camara, C., Putterman, S. & Kirilov, E. Sonoluminescence from a single bubble driven at 1 megahertz. *Phys. Rev. Lett.* **92**, 124301 (2004).
- Yasui, K. Single-bubble sonoluminescence from noble gases. *Phys. Rev. E* **63**, 035301 (2001).
- Ajello, J. M. *et al.* Middle ultraviolet and visible spectrum of SO_2 by electron impact. *J. Geophys. Res. Space* **107**, SIA2 (2002).
- Schappe, R. S., Schulman, M. B., Sharpton, F. A. & Lin, C. C. Emission of the O_2^+ ($A^2\Pi_u \rightarrow X^2\Pi_g$) second-negative-band system produced by electron impact on O_2 . *Phys. Rev. A* **38**, 4537–4545 (1988).

Acknowledgements This work was supported by the National Science Foundation and the US Defense Advanced Research Projects Agency. We acknowledge conversations with F. Grieser on the mechanism of Ar atom emission, and with L. A. Crum, D. Lohse, W. C. Moss and S. J. Putterman.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.S.S. (ksuslick@uiuc.edu).

Self-directed self-assembly of nanoparticle/copolymer mixtures

Yao Lin^{1*}, Alexander Böker^{1*†}, Jinbo He¹, Kevin Sill¹, Hongqi Xiang¹, Clarissa Abetz², Xuefa Li³, Jin Wang³, Todd Emrick¹, Su Long⁴, Qian Wang⁴, Anna Balazs⁵ & Thomas P. Russell¹

¹Department of Polymer Science & Engineering, University of Massachusetts, Amherst, Massachusetts 01003, USA

²Bayreuther Institut für Makromolekülforschung, Universität Bayreuth, 95440 Bayreuth, Germany

³Advanced Photon Source, Argonne National Laboratory, Argonne, Illinois 60439, USA

⁴Department of Chemistry, University of South Carolina, Columbia, South Carolina 29208, USA

⁵Department of Chemical and Petroleum Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA

* These authors contributed equally to this work

† Present address: Lehrstuhl für Physikalische Chemie II, Universität Bayreuth, 95440 Bayreuth, Germany

The organization of inorganic nanostructures within self-assembled organic or biological templates^{1–11} is receiving the attention of scientists interested in developing functional hybrid materials. Previous efforts have concentrated on using such scaffolds^{7,9,12} to spatially arrange nanoscopic elements as a strategy for tailoring the electrical, magnetic or photonic properties^{8–11,13–16} of the material. Recent theoretical arguments^{16–18} have suggested that synergistic interactions between self-organizing particles and a self-assembling matrix material can lead to hierarchically ordered structures. Here we show that mixtures of diblock copolymers and either cadmium selenide- or ferritin-based nanoparticles exhibit cooperative, coupled self-assembly on the nanoscale. In thin films, the copolymers assemble into cylindrical domains, which dictate the spatial distribution of the nanoparticles; segregation of the particles to the interfaces mediates interfacial interactions and orients the copolymer domains normal to the surface, even when one of the blocks is strongly attracted to the substrate. Organization of both the polymeric and particulate entities is thus achieved without the use of external fields^{10,19}, opening a simple and general route for fabrication of nanostructured materials with hierarchical order.

Block copolymer/nanoparticle films were prepared by spin-coating toluene solutions of a mixture of 3- or 5-wt% polystyrene-*block*-poly(2-vinylpyridine) copolymer, denoted PS-*b*-P2VP, and 1-wt% tri-*n*-octylphosphine oxide-(TOPO)-covered CdSe nanoparticles (4 nm in diameter) onto silicon wafers. The 150–600-nm-thick films were annealed thermally at 170 °C under vacuum, in a supercritical fluid CO₂ environment at 70 °C, or in chloroform vapours at room temperature (see Supplementary Information). Each treatment imparts mobility to the thin film mixtures, allowing them to attain an equilibrium morphology within about two days. We note that the TOPO-covered CdSe nanoparticles are stable at the temperatures used, such that aggregation caused by ligand disassociation was not observed. Films of the pure diblock copolymer and the block copolymer mixed with TOPO were prepared and used as controls to assess the influence of the nanoparticles on the thin-film behaviour.

Figure 1a and b show scanning force microscopy (SFM) height and phase images of a 400-nm-thick PS-*b*-P2VP film spin-coated from toluene and dried. An array of microphase-separated domains of P2VP in a PS matrix, similar to those found in previous investigations^{20,21}, is seen. The phase image maps onto the height variations from the topography of the sample. Upon thermal annealing at 170 °C for two days, the equilibrium morphology of

the polymer—that is, cylindrical microdomains of P2VP in a PS matrix—develops. All surface features are lost upon annealing, as seen in Fig. 1c and d. We note here that the control samples composed of mixtures of the block copolymer and TOPO showed the same behaviour. The preferential interaction of P2VP with the substrate and the lower surface energy of PS force an orientation of the cylindrical microdomains parallel to the substrate. Incommensurability between the film thickness and period of the copolymer leads to the formation of islands and holes on the surface with a step height of one period. These features are macroscopic in size and the copolymer film is otherwise smooth; structural features on the nanoscopic level are absent^{22–24}.

With the block copolymer–nanoparticle mixtures, a completely different behaviour is observed. Figure 1e and f show the SFM height and phase images of the as-spun sample. An array of microphase-separated domains, identical to that of the pure PS-*b*-P2VP, can be seen. However, after thermal annealing (Fig. 1g, h), hexagonally ordered features are found on the film surface. The very pronounced phase shift indicates that the hard nanoparticles have been incorporated into the P2VP phase. The hexagonal packing suggests that the addition of nanoparticles has caused an orientation of the cylindrical P2VP microdomains normal to the substrate. From Fourier transformation of the SFM images, a change in the lattice spacing of the cylindrical domains from 320 ± 20 to 480 ± 20 Å is seen upon annealing—this is consistent with SAXS data obtained from the bulk samples. This indicates that the system has reached equilibrium. Similar surface features were also found for films that were annealed in supercritical fluid CO₂ or chloroform vapour. The SFM images from these latter experiments are shown in Fig. 1i and j. Depth profiles obtained by sequential SEM images of oxygen-plasma-etched samples showed a hexagonal array of microdomains oriented normal to the film surface that persisted throughout the film (see Supplementary Information). The above results indicate that, given sufficient mobility, the microdomains reorient to form hexagonally packed cylindrical microdomains oriented normal to the film surface and substrate interface. This result was found to be independent of film thickness. Consistent with the results of Kim *et al.*, enhanced lateral ordering was observed upon solvent-vapour annealing²⁵.

To understand better the morphological changes in the thin-film blend, we carried out grazing incidence small-angle X-ray scattering (GISAXS) measurements before (Fig. 2a) and after (Fig. 2b–d) thermal annealing. After spin-coating, a weak but readily observable in-plane scattering peak at 0.0196 Å^{-1} indicates the initial ordering of the diblock copolymer with a *d*-spacing of 320 Å. The average separation distance between the particles is too large to be observed (inset to Fig. 2a). Upon annealing at 170 °C, the nanoparticles preferentially segregate to the cylindrical P2VP microdomains where the lateral ordering of the microdomains has improved. Consequently, two orders of the in-plane diffraction peaks are seen at 0.0131 and 0.0227 Å^{-1} , corresponding to the (10) and (11) reflections of a two-dimensional hexagonal lattice with a *d*-spacing of 480 Å. This is consistent with the structure observed in the bulk by SAXS.

The X-ray penetration depth can be controlled by the incident angle, providing details on the depth dependence of the thin-film morphology. A series of incident angles ranging from 0.04° to 0.12° (the critical angle, θ_c , was measured to be 0.11°) were used to achieve penetration depths from 44 Å to full penetration into the film. Below θ_c (for example, 0.08° and 0.09°), as shown in Fig. 2b and c, when the X-ray penetration depth is 54 and 61 Å, respectively, the (10) and (11) scattering peaks are clearly observable. As soon as the incident angle is slightly above θ_c , only the (10) reflection is pronounced (shown in Fig. 2d at 0.12°), suggesting a slight reduction in the hexagonal ordering and scattering contrast of the microdomains due to a decrease in the nanoparticle concentration within the film.

This is seen more clearly in Fig. 2e, where the q_y linescans at $q_z = 0.0376 \text{ Å}^{-1}$ (q_y and q_z being the momentum transfer normal to the diffraction plane and normal to the sample surface, respectively) are shown at four incident angles, corresponding to

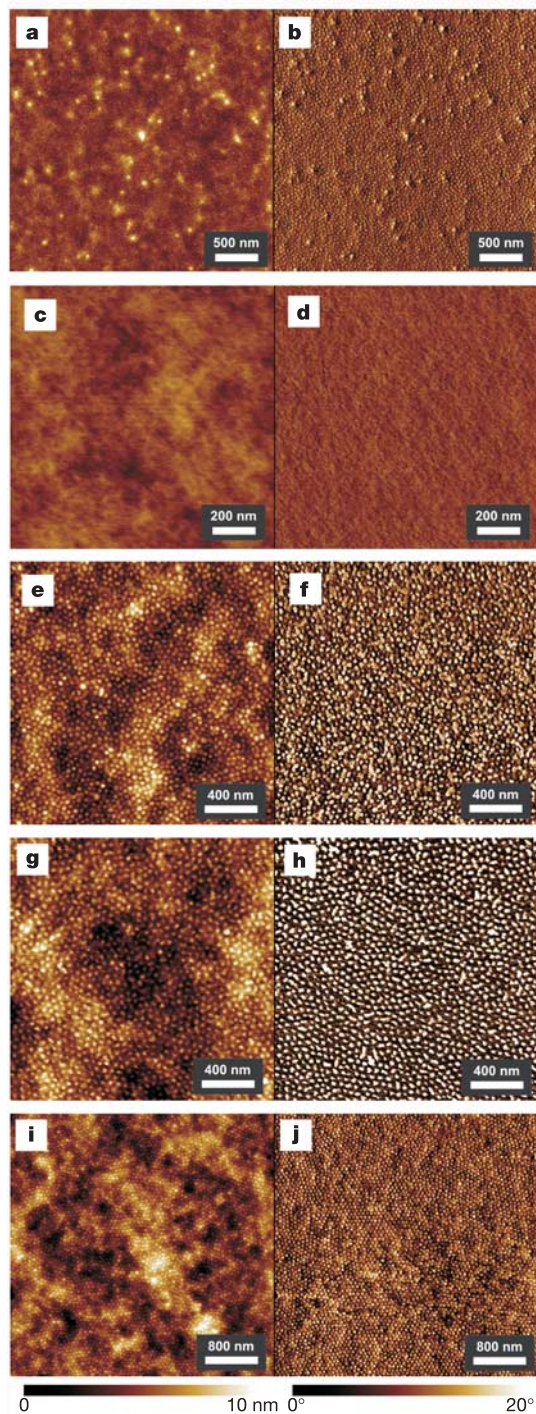


Figure 1 SFM topography and phase images of thin films from pure PS-*b*-P2VP block copolymer and PS-*b*-P2VP–CdSe nanoparticle composites. **a–d**, SFM topography (**a**, **c**) and phase images (**b**, **d**) of a PS-*b*-P2VP block copolymer film taken after spin-coating (**a**, **b**) and after thermal annealing at 170 °C for 2 days (**c**, **d**). **e–j**, SFM topography (**e**, **g**, **i**) and phase images (**f**, **h**, **j**) of films prepared from a mixture of PS-*b*-P2VP block copolymer and CdSe nanoparticles. **e**, **f**, Structure of an as-spun film. **g**, **h**, A film after thermal annealing at 170 °C for 2 days. **i**, **j**, A film after annealing in saturated chloroform solvent vapour for 1 day. Z-range: 10 nm, 20°.

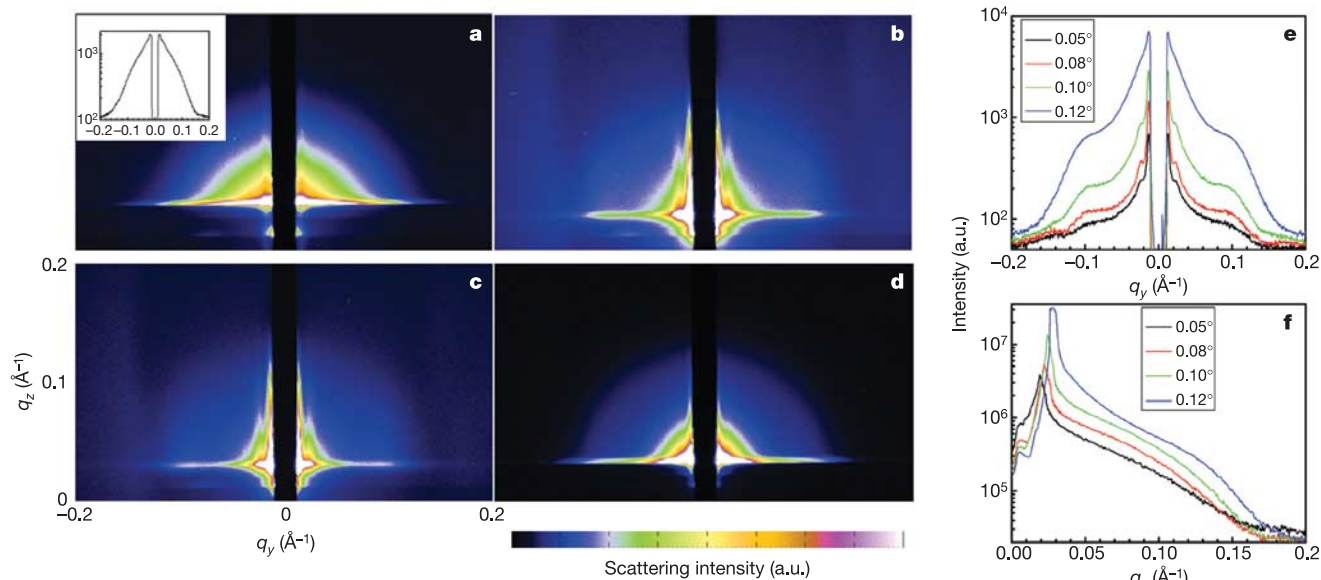


Figure 2 GISAXS data on the thin film of PS-*b*-P2VP-CdSe nanoparticle composites. Experimental data from GISAXS measurements. **a**, Data from a freshly spin-coated thin film of the diblock copolymer and nanoparticles at an incident angle of 0.06° with a penetration depth of $44 \text{ \AA}$. **b**, Data from a thin film (0.08° , $54 \text{ \AA}$) thermally annealed for 2 days at 170°C . **c**, Data from the same annealed film (0.09° , $61 \text{ \AA}$). **d**, Data from the same annealed film (0.12° , full thickness). **e**, q_y linescan at $q_z = 0.0376 \text{ \AA}^{-1}$ at four

different X-ray incident angles of 0.05° , 0.08° , 0.10° and 0.12° , with penetration depths of 44 , 54 , $73 \text{ \AA}$ and full thickness, respectively. **f**, q_z linescan along the first-order peak ($1\ 0$) for the four incident angles. The critical angle of the thin film is measured to be 0.11° and q_y and q_z are the momentum transfers normal to the diffraction plane and normal to the sample surface.

penetration depths of 44 , 54 and $73 \text{ \AA}$, and the full film thickness, respectively. In addition, the spatial correlation of the nanoparticles is characterized by the 'shoulder' in the data at $q_y \approx 0.1 \text{ \AA}^{-1}$, corresponding to an average distance between the nanoparticles of $\sim 60 \text{ \AA}$, as would be expected for a closely packed nanoparticle array. We note that the bulk concentration of the nanoparticles is low and the average interparticle distance is too large to be seen by GISAXS. The nanoparticles are closely packed and the scattering 'shoulder' near $q_y = 0.1 \text{ \AA}^{-1}$ can be seen, even at very low incident angles where the penetration depth is only a fraction of the nanoparticle size, so the nanoparticles must be concentrated near the air-film interface. Figure 2f shows linescans as a function of the momentum transfer at $q_y = 0.0131 \text{ \AA}^{-1}$ (the $(1\ 0)$ plane) for different incident angles. The curves in Fig. 2f can be fitted to obtain the spatial location of the nanoparticles in the P2VP cylinders quantitatively. A detailed analysis is provided in the Supplementary Information. Together with the SFM data, these results indicate that the nanoparticles cap the P2VP cylinders at the air surface. GISAXS investigations of the solvent-annealed samples yielded similar results.

Whereas GISAXS provides an average orientation of the microdomains in the film, TEM was used to determine whether the orientation extended from the substrate to the film surface (see Supplementary Information for experimental methods). A cross-sectional TEM image of pure PS-*b*-P2VP film after annealing is shown in Fig. 3a. Here, all the cylindrical microdomains are oriented parallel to the substrate. Figure 3b shows a TEM image of a PS-*b*-P2VP film mixed with nanoparticles after thermal annealing. These data quantitatively show the penetration of the microdomains normal to the surface, and the persistence of order over very large distances. Owing to the low magnification necessary to observe the morphology over large distances, individual particles cannot be discerned. Unstained samples reveal that the nanoparticles preferentially reside within the P2VP microdomains. The higher-resolution SEM image in Fig. 3c, a top view of a

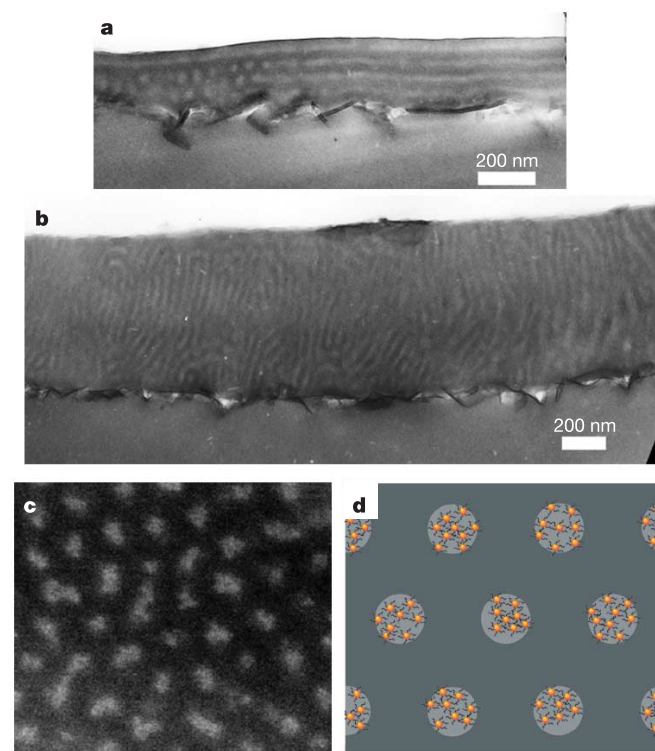


Figure 3 TEM and SEM images of thin films from pure PS-*b*-P2VP block copolymer and PS-*b*-P2VP-CdSe nanoparticle composites. **a**, TEM image of cross-section of a pure PS-*b*-P2VP block copolymer film after annealing at 170°C for 2 days. **b**, TEM image of cross-section of a PS-*b*-P2VP block copolymer-CdSe nanoparticle film after annealing at 170°C for 2 days. **c**, Secondary electron SEM image of the surface of a thin film of a block copolymer-CdSe nanoparticle mixture spin-coated onto a silicon wafer and annealed at 170°C for 2 days (image width, 250 nm) taken at 1-kV acceleration voltage. **d**, Schematic representation of nanoparticle assembly at the P2VP cylinders.

PS-*b*-P2VP–nanoparticle mixture film, shows that the nanoparticles also assemble on the surface of the P2VP cylinders. Although there is only very weak contrast between the PS and P2VP, the very bright areas are attributed to the CdSe nanoparticles, almost completely capping the tops of the P2VP cylindrical microdomains (schematically shown in Fig. 3d). Thus, the cylindrical microdomains of the higher-surface-energy P2VP ($\gamma_{\text{P2VP}} \approx 47 \text{ mN m}^{-1}$)²⁶ are coated with the lower-surface-energy, hydrocarbon-coated CdSe nanoparticles ($\gamma_{\text{hydrocarbon}} = 30\text{--}33 \text{ mN m}^{-1}$)²⁶, effectively balancing the surface interactions relative to the PS matrix ($\gamma_{\text{PS}} \approx 39 \text{ mN m}^{-1}$)²⁶. This causes an orientation of the cylindrical microdomains normal to the surface. Experiments on different substrates and film thicknesses showed that the perpendicular orientation is found regardless of the nature of the substrate or the film thickness. Consequently, we deduce that interfacial interactions with the substrate are mediated in a manner similar to that at the air surface.

This interplay between assembly processes should be applicable to a wide variety of other systems, because the surface chemistry of the nanoparticles can be tuned. To test this, we used a blend of poly(ethylene glycol) (PEG)-tagged ferritin bio-nanoparticles, denoted ferritin-PEG (see Supplementary Information), and a lamella-forming diblock copolymer of P2VP and poly(ethylene oxide), denoted P2VP-*b*-PEO. Thin-film samples of the block copolymer with and without ferritin-PEG particles were prepared and annealed in saturated benzene vapour. Without ferritin-PEG particles, the P2VP-*b*-PEO microphase-separated into lamellae oriented parallel to the surface with the crystalline PEO located at the surface, as seen by optical microscopy (Fig. 4a) and SFM (Fig. 4b and 4c). With ferritin-PEG, the PEO crystallization is suppressed

(Fig. 4d), and the lamellar microdomains orient normal to the surface (Fig. 4e and f). Thus, the ferritin-PEG bio-nanoparticles are incorporated into PEO microdomains, suppress crystallization, mediate interfacial interactions and reorient the microdomains. This reorientation is similar to that seen in the PS-*b*-P2VP–CdSe nanoparticle mixtures. With this ability to genetically and chemically^{27,28} manipulate the surface properties of bio-nanoparticles such as ferritin, and to incorporate different inorganic materials into the cores using biomineralization, copolymer–bioparticle hybrid systems can be developed with unique functionalities.

We have shown here examples of synthetic and biologically inspired systems, where a one-step hierarchical self-organization occurs via an interplay between distinct self-assembling processes, producing spatially ordered, organic–inorganic and organic–bioparticle hybrid materials. This synergy represents a significant advance over other processes that rely on sequential fabrication steps to incorporate functionality into pre-organized templates^{1–4,10,11}. Furthermore, the orientation of microdomains, built into the system by the segregation of the nanoparticles to the interfaces, is independent of film thickness, the nature of the interfaces and the geometry of the system and eliminates the need for external fields to manipulate the orientation. This is particularly important for systems like PS-*b*-P2VP, where interfacial interactions are so strong that even large external fields cannot reorient the domains through the entire film. Coupled with recent advances in the synthesis of nanocrystals²⁹ and the large variety of bionanoparticles that can be surface-modified²⁸ and biomineralized³⁰, the synergistic assembly process described here provides remarkable control and flexibility over the fabrication of nanostructured

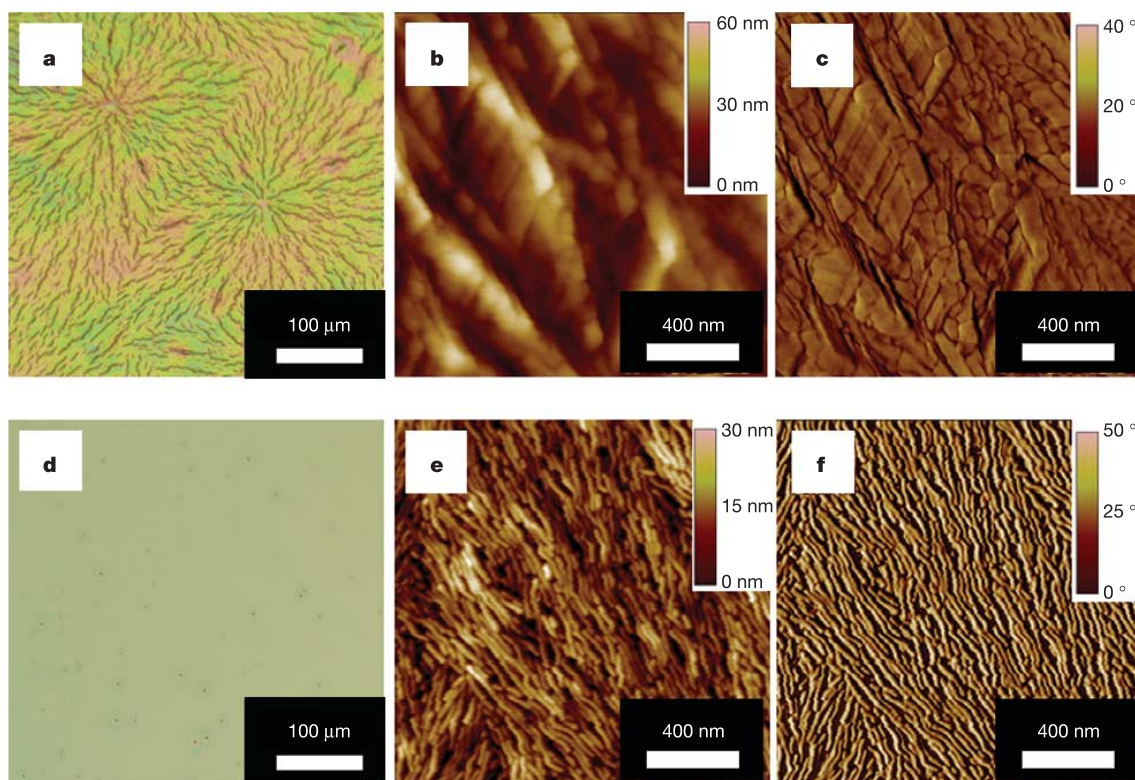


Figure 4 Optical microscopy and SFM images of thin films from pure P2VP-*b*-PEO block copolymer and P2VP-*b*-PEO–ferritin-PEG nanoparticle composites. Optical microscopy images (**a**, **d**) and SFM topography (**b**, **e**) and phase images (**c**, **f**) of a P2VP-*b*-PEO thin film after annealing in saturated benzene vapour for 3 days with and without ferritin-PEG.

a–c, Structure of a P2VP-*b*-PEO thin film after annealing, without ferritin-PEG nanoparticles. **d–f**, Structure of a P2VP-*b*-PEO thin film after annealing, with ferritin-PEG nanoparticles.

materials with applications including chemical sensing, separation, catalysis, high-density data storage and photonic materials. □

Received 14 October; accepted 21 December 2004; doi:10.1038/nature03310.

1. Spatz, J. P. *et al.* Ordered deposition of inorganic clusters from micellar block copolymer films. *Langmuir* **16**, 407–415 (2000).
2. Lopes, W. A. & Jaeger, H. M. Hierarchical self-assembly of metal nanostructures on diblock copolymer scaffolds. *Nature* **414**, 735–738 (2001).
3. Pai, R. A. *et al.* Mesoporous silicates prepared using preorganized templates in supercritical fluids. *Science* **303**, 507–510 (2004).
4. Templin, M. *et al.* Organically modified aluminosilicate mesostructures from block copolymer phases. *Science* **278**, 1795–1798 (1997).
5. Liang, H. J., Angelini, T. E., Ho, J., Braun, P. V. & Wong, G. C. L. Molecular imprinting of biomimeticized CdS nanostructures: Crystallographic control using self-assembled DNA-membrane templates. *J. Am. Chem. Soc.* **125**, 11786–11787 (2003).
6. Li, M., Schnablegger, H. & Mann, S. Coupled synthesis and self-assembly of nanoparticles to give structures with controlled organization. *Nature* **402**, 393–395 (1999).
7. Bockstaller, M. R., Lapetnikov, Y., Margel, S. & Thomas, E. L. Size-selective organization of enthalpic compatibilized nanocrystals in ternary block copolymer/particle mixtures. *J. Am. Chem. Soc.* **125**, 5276–5277 (2003).
8. Shenton, W., Pum, D., Sleytr, U. B. & Mann, S. Synthesis of cadmium sulphide superlattices using self-assembled bacterial S-layers. *Nature* **389**, 585–587 (1997).
9. Mann, S. & Ozin, G. A. Synthesis of inorganic materials with complex form. *Nature* **382**, 313–318 (1996).
10. Thurn-Albrecht, T. *et al.* Ultrahigh-density nanowire arrays grown in self-assembled diblock copolymer templates. *Science* **290**, 2126–2129 (2000).
11. Bockstaller, M., Kolb, R. & Thomas, E. L. Metallo-dielectric photonic crystals based on diblock copolymers. *Adv. Mater.* **13**, 1783–1786 (2001).
12. Black, C. T., Murray, C. B., Sandstrom, R. L. & Sun, S. H. Spin-dependent tunneling in self-assembled cobalt-nanocrystal superlattices. *Science* **290**, 1131–1134 (2000).
13. Sanchez, C. & Lebeau, B. Design and properties of hybrid organic-inorganic nanocomposites for photonics. *MRS Bull.* **26**, 377–387 (2001).
14. Alexandre, M. & Dubois, P. Polymer-layered silicate nanocomposites: preparation, properties and uses of a new class of materials. *Mater. Sci. Eng. Rev.* **28**, 1–63 (2000).
15. Soo, P. P. *et al.* Rubbery block copolymer electrolytes for solid-state rechargeable lithium batteries. *J. Electrochem. Soc.* **146**, 32–37 (1999).
16. Balazs, A. C. Interactions of nanoscopic particles with phase-separating polymeric mixtures. *Curr. Opin. Colloid Interf. Sci.* **4**, 443–448 (2000).
17. Lee, J. Y., Shou, Z. & Balazs, A. C. Modeling the self-assembly of copolymer-nanoparticle mixtures confined between solid surfaces. *Phys. Rev. Lett.* **91**, 136103 (2003).
18. Lee, J. Y., Shou, Z. & Balazs, A. C. Predicting the morphologies of confined copolymer/nanoparticle mixtures. *Macromolecules* **36**, 7730–7739 (2003).
19. Mansky, P., Liu, Y., Huang, E., Russell, T. P. & Hawker, C. Controlling polymer-surface interactions with random copolymer brushes. *Science* **275**, 1458–1460 (1997).
20. Meiners, J. C., Quintel-Ritzi, A., Mlynek, J., Elbs, H. & Krausch, G. Adsorption of block-copolymer micelles from a selective solvent. *Macromolecules* **30**, 4945–4951 (1997).
21. Li, Z. *et al.* Self-ordering of diblock copolymers from solution. *J. Am. Chem. Soc.* **118**, 10892–10893 (1996).
22. Liu, Y. *et al.* Surface-induced ordering in asymmetric block copolymers. *Macromolecules* **27**, 4000–4010 (1994).
23. Morkved, T. L. & Jaeger, H. M. Thickness-induced morphology changes in lamellar diblock copolymer ultrathin films. *Europhys. Lett.* **40**, 643–648 (1997).
24. Knoll, A., Magerle, R. & Krausch, G. Phase behavior in thin films of cylinder-forming ABA block copolymers: Experiments. *J. Chem. Phys.* **120**, 1105–1116 (2004).
25. Kim, S. H., Misner, M. J., Xu, T., Kimura, M. & Russell, T. P. Highly oriented and ordered arrays from block copolymers via solvent evaporation. *Adv. Mater.* **16**, 226–231 (2004).
26. Brandrup, J. & Immergut, E. H. *Polymer Handbook* 3rd edn, VI/411–VI/434 (John Wiley & Sons, New York, 1998).
27. Wong, K. K. W., Whilton, N. T., Colfen, H., Douglas, T. & Mann, S. Hydrophobic proteins: synthesis and characterisation of organic-soluble alkylated ferritins. *Chem. Commun.*, 1621–1622 (1998).
28. Raja, K. S. & Wang, Q. *Encyclopedia of Nanoscience and Nanotechnology* 321–330 (Marcel Dekker, New York, 2004).
29. Alivisatos, A. P. Perspectives on the physical chemistry of semiconductor nanocrystals. *J. Phys. Chem.* **100**, 13226–13239 (1996).
30. Meldrum, F. C., Wade, V. J., Nimmo, D. L., Heywood, B. R. & Mann, S. Synthesis of inorganic nanophase materials in supramolecular protein cages. *Nature* **349**, 684–687 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Xu, J. Sievert, K. Lavery, K. Schmidt, M. Hund, X. Hu, H. Skaff and J. Hirsch for their assistance with the experiments. This work was supported by the US Department of Energy, the NSF-supported MRSEC at the University of Massachusetts Amherst, the NSF Collaborative Research in Chemistry Program, the NSF Career Award, the Army Research Office through a MURI, the NSF CRC programme, and the MAX KADE Foundation. The Advanced Photon Source is supported by the US Department of Energy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to TPR (russell@mail.pse.umass.edu).

Mesozoic Alpine facies deposition as a result of past latitudinal plate motion

Giovanni Muttoni^{1,2}, Elisabetta Erba¹, Dennis V. Kent³ & Valerian Bachtadse⁴

¹Department of Earth Sciences, University of Milan, via Mangiagalli 34, I-20133 Milan, Italy

²ALP – Alpine Laboratory of Paleomagnetism, via Madonna dei Boschi 76, I-12016 Peveragno (CN), Italy

³Department of Geological Sciences, Rutgers University, Piscataway, New Jersey 08854, USA, and Lamont-Doherty Earth Observatory, Palisades, New York 10964, USA

⁴Department fuer Geo- und Umweltwissenschaften, Ludwig-Maximilians-Universitaet Muenchen, Theresienstrasse 41, D-80333 Munich, Germany

The fragmentation of Pangaea as a consequence of the opening of the Atlantic Ocean is documented in the Alpine–Mediterranean region by the onset of widespread pelagic sedimentation¹. Shallow-water sediments were replaced by mainly pelagic limestones in the Early Jurassic period, radiolarian cherts in the Middle–Late Jurassic period, and again pelagic limestones in the Late Jurassic–Cretaceous period¹. During initial extension, basin subsidence below the carbonate compensation depth (CCD) is thought to have triggered the transition from Early Jurassic limestones to Middle–Late Jurassic radiolarites¹. It has been proposed that the transition from radiolarites to limestones in the Late Jurassic period was due to an increase in calcareous nannoplankton abundance when the CCD was depressed below the ocean floor¹. But in modern oceans, sediments below the CCD are not necessarily radiolaritic. Here we present palaeomagnetic samples from the Jurassic–Cretaceous pelagic succession exposed in the Lombardian basin, Italy. On the basis of an analysis of our palaeolatitudinal data in a broader palaeogeographic context, we propose an alternative explanation for the above facies tripartition. We suggest that the Lombardian basin drifted initially towards, and subsequently away from, a near-equatorial upwelling zone of high biosiliceous productivity. Our tectonic model for the genesis of radiolarites adds an essential horizontal plate motion component to explanations involving only vertical variations of CCD relative to the ocean floor. It may explain the deposition of radiolarites throughout the Mediterranean and Middle Eastern region during the Jurassic period.

Modern oceans are undersaturated with respect to silica², and this causes the post-mortem dissolution of siliceous tests. Only under specific oceanic conditions, whereby the wind-driven upwelling of nutrient- and silica-rich deeper waters sustain enhanced plankton productivity, can a significant fraction of the siliceous biomass be buried before dissolution. Radiolarians in particular tend to proliferate in the equatorial belts of the Pacific and Indian oceans². The hypothesis that Mesozoic Alpine facies changes resulted from past latitudinal plate motion across climate belts can be tested with palaeomagnetism. Rocks can acquire a magnetization parallel to the Earth's ambient geomagnetic field at the time of their formation. Because the time-averaged field is essentially dipolar, the inclination of the natural remanent magnetization with respect to the palaeo-horizontal, represented by sedimentary bedding planes, is a function of palaeolatitude. We studied the palaeomagnetism of the Jurassic–Cretaceous pelagic succession exposed at Colle Sogno in the Lombardian basin (Fig. 1); this location was selected because of its relatively simple tectonic setting, quality of exposure and the stratigraphic continuity of strata of known age. The succession, biostratigraphically dated with nannofossils and ammonites (see

refs 3 and 4, and references therein), was assigned numerical ages according to the timescale of ref. 5 (Fig. 2). The lower part of the succession consists of 450 m of shallow water limestones (Zu Formation–Conchodon Dolomite), 550–600 m of grey calcarenites and calcilutites with chert lenses and marly interbeds (Sedrina, Moltrasio and Domaro Formations) and 145 m of grey-pink pelagic limestones and marlstones (Sogno Formation), collectively of Rhaetian–early Bajocian age. The middle part consists of 20 m of ribbon-bedded green radiolarites, 15 m of red knobby radiolarites and 29 m of red cherty limestones (Rosso ad Aptici Formation), collectively of early Bajocian–late Tithonian age. The upper part consists of 150 m of nannofossil-rich calcilutites of the Maiolica Formation (Tithonian–Aptian). Sediment accumulation rates (not corrected for dewatering and burial compaction) are estimated at $\sim 30 \text{ m Myr}^{-1}$ in the Moltrasio–Domaro Formations, 12 m Myr^{-1} in the Sogno Formation, $2\text{--}3 \text{ m Myr}^{-1}$ in the radiolarites–Rosso ad Aptici interval, and increasing to $6\text{--}7 \text{ m Myr}^{-1}$ in the Maiolica Formation (Fig. 2).

Oriented palaeomagnetic samples (191) were collected at 13 sites through the succession (Fig. 2, sites VC and CS1–CS11). Samples were thermally demagnetized (Supplementary Table A) in order to retrieve characteristic magnetization components (Supplementary Table B) and to calculate site-mean palaeomagnetic poles and palaeolatitudes with associated errors (Supplementary Table C) (see Methods). The time span covered by each sampling site ($\sim 0.1\text{--}1.5 \text{ Myr}$) should be sufficient to average out secular variations. Literature data from the Domaro⁶ and Maiolica^{7–9} Formations of the Lombardian basin, projected to the Colle Sogno locality coordinates, augmented the definition of our palaeolatitude curve, and extended it into the Early Cretaceous.

The Lombardian basin composite palaeolatitude curve (Fig. 2; Supplementary Table C) shows a moderate southward drift from 35° N to 25° N in the Early Jurassic followed by a drop to values of 10° N and less in the Middle–Late Jurassic coincident with

radiolarite deposition (sites CS8–CS10). Near-equatorial latitudes during Middle–Late Jurassic deposition of radiolarites were also palaeomagnetically recorded outside the Lombardian basin in Tuscany¹⁰ (Fig. 1). The mean palaeopole from these radiolarites is offset by $\sim 70^\circ$ anticlockwise from site CS8 of this study. This offset is due to vertical axis rotation with negligible latitudinal transport. Accordingly, we can incorporate data from the Tuscan radiolarites in the Lombardian basin palaeolatitude curve (TU1 and TU2; Fig. 2). Higher values of $20\text{--}30^\circ \text{ N}$ occurred in the Early Cretaceous during deposition of the Maiolica nannofossil ooze.

Sedimentary compaction may cause the palaeomagnetic inclination of some sediment to be shallower than the true inclination of the palaeomagnetic field. To check this possibility, palaeolatitudes from well-dated magmatic rocks from Africa and North America were rotated to northwest Africa coordinates using published Euler poles, and transferred to the Colle Sogno locality (Supplementary Table C). This transfer procedure is allowed by the palaeogeographic affinity of Adria and Africa, as demonstrated by the fact that the northern margin of Adria in the Lombardian basin and elsewhere in the Southern Alps, although locally deformed during the Alpine orogeny, has Permian–Mesozoic palaeomagnetic directions that are statistically indistinguishable from those of Africa^{11,12}. Tropical latitudes for the Lombardian basin during the Early Jurassic are corroborated by data from magmatic rocks from Sierra Leone and Mauritania (data points FR¹³, NM and SM¹⁴ in Fig. 2). Near-equatorial latitudes in the Middle–Late Jurassic are supported by data from the Mzongwana kimberlite of South Africa dated at $152 \pm 3.4 \text{ Myr}$ ago (MZ¹⁵). More tropical latitudes in the Early Cretaceous are confirmed by data from the Swarttruggens Main kimberlite of South Africa (SMA¹⁵), the Ithaca kimberlites (NY¹⁶) of North America, the White Mountains plutons (NE¹⁷) of North America, and the Group 2 kimberlites of South Africa (G2K¹⁵) (Fig. 2; Supplementary Table C).

We placed the palaeolatitudinal evolution in a broader palaeogeographic perspective by calculating overall mean palaeopoles for Adria (Supplementary Table C) and restoring other Gondwanan and Laurasian elements relative to northwest Africa (Adria) using published Euler poles of rotation (Fig. 2, right panels) (see Methods). The rapid southward movement of Adria/Africa from the Early Jurassic (187–177 Myr ago) to the Middle–Late Jurassic (166–149 Myr ago) was broadly coeval with the opening of a seaway connecting the central Atlantic Ocean to the Tethys Ocean, which around Adria is called the Liguria–Piedmont Ocean. Inclination-only palaeomagnetic data from the Middle East support our 166–149 Myr reconstruction by indicating that the northern margin of Arabia was close to the Equator in the Late Jurassic–Early Cretaceous¹⁸. The late Middle Jurassic to Late Jurassic was the time of maximum abundance of radiolarian-rich sediments in the Atlantic–Tethys seaway and surrounding margins at sites in Spain, the Alps, Carpathians, Adria and Greece, as well as in the Tethys Ocean facing Oman^{19,20}. These sites were located at near-equatorial latitudes, presumably in a zone of high biosiliceous productivity sustained by upwelling of deeper waters enriched in nutrients and silica^{19,20}, as observed in modern equatorial belts². This is confirmed by a recent study of Jurassic Apennines radiolarites that indicates repeated periods of high primary productivity of surface waters sustained by nutrients remobilized by strong ocean circulation²¹. In the time span when the Atlantic–Tethys seaway resided at near-equatorial latitudes, the actual deposition of radiolarites depended also on local or regional palaeogeographic and/or palaeoecologic factors. Low fertility related to dominant antiestuarine-type circulation, relatively shallow water depth promoting preservation of carbonates, dilution by terrigenous input, or a combination of these factors, may explain the absence of true radiolarites in the young, land-locked subequatorial central Atlantic Ocean^{e.g.,20}. The upper Middle–lower Upper Jurassic sections at Deep Sea Drilling Project Sites 367 and 534 consist in any case of radiolarian-rich intervals

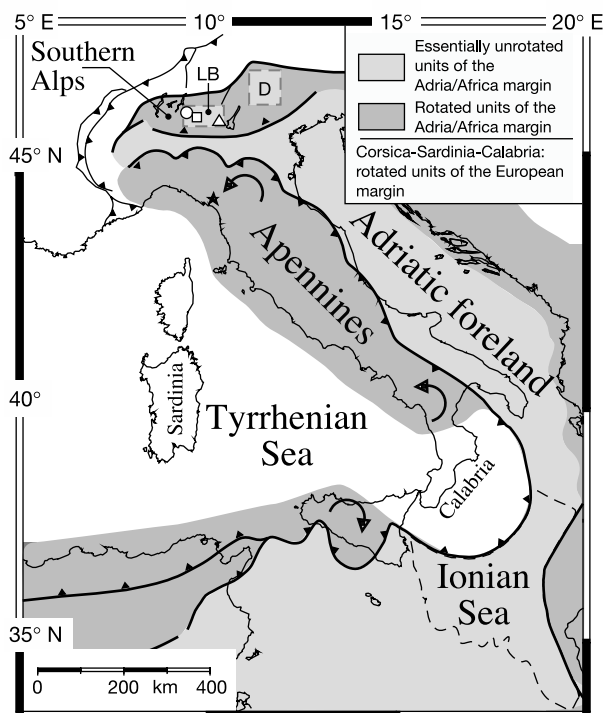


Figure 1 Structural map of Italy. Key palaeomagnetic sites from the Southern Alps and the Apennines discussed in the text are: open square, Valcava-Colle Sogno; open circle, Alpetto⁷; open triangle, Piè del Dosso⁸; filled star, Tuscany¹⁰. D, Dolomites, LB, Lombardian basin. Site coordinates are in Supplementary Table C.

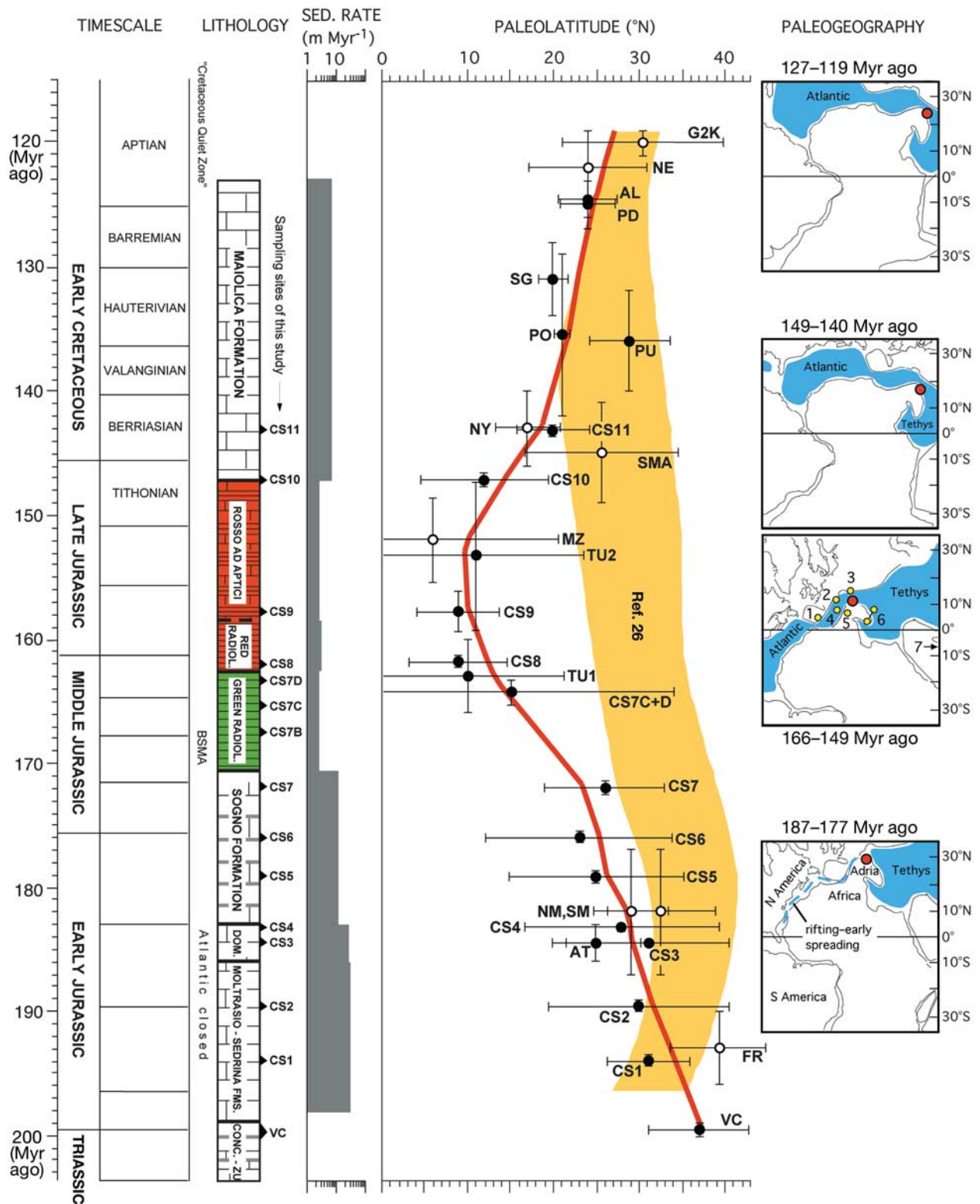


Figure 2 Palaeolatitudinal evolution of Colle Sogno. Left panels, timescale of ref. 5 with marine magnetic anomalies (M37–M0r; BSMA, Blake Spur) placed next to the Colle Sogno succession (CONC., Conchodon Dolomite; ZU, Zu Formation; DOM., Domaro Formation; RADIOL., radiolarites) with average sedimentation (SED.) rates. Central panels: palaeolatitudes from sediments (filled circles) and magmatic rocks (open circles; acronyms in Supplementary Table C) interpolated by the red line compared with palaeolatitudes from Besse and Courtillot²⁶ (yellow band). Right panels: Adria's palaeogeography (oceanic crust in blue; see Methods); yellow circles: Spain (1), Swiss

Alps (2), Carpathians (3), western Alps–Liguria (4), Tuscany–Umbria (5), Greece (6), Oman (7); Lombardian basin: red circle. Horizontal error bars represent paleolatitude uncertainty expressed by $\pm A95$ (Fisher's cone of 95% confidence about the paleomagnetic pole) or approximated by $\pm dp$ (minor axis of Fisher's 95% confidence oval about the paleomagnetic pole) (see Methods and Supplementary Table C). Vertical error bars represent age coverage of palaeomagnetic sites derived from sediment accumulation rates estimates (see text).

that suggest fertility regimes higher than in preceding and following intervals²⁰.

Radiolarites deposited directly on exhumed serpentinized mantle rocks and pillow lavas are a key element of Alpine ophiolites in the so-called "Steinmann trinity"²². Oceanic gabbros from Corsica, Liguria and the eastern Alps occur in a narrow age window centred at 160 Myr ago (see references in ref. 22) when the Liguria–Piedmont Ocean was at near-equatorial latitudes. This would explain why radiolarites rather than carbonates were deposited on the newly formed and thus presumably shallow oceanic crust.

Radiolarian cherts deposited in equatorial zones of ancient oceans are a diachronous facies controlled primarily by plate motion across oceanic zonal circulation patterns. Cherts deposited at near-equatorial palaeolatitudes in the Slide Mountain displaced terrane of British Columbia are late Palaeozoic in age²³, and those of the Franciscan Complex accreted on the continental margin of California and Mexico are Jurassic in age²⁴, whereas those deposited in response to the equatorial crossing of the Pacific plate are Late Jurassic and Cretaceous²⁵. At Colle Sogno, oceanic sedimentation became progressively less siliceous and more calcareous in the Early Cretaceous (Fig. 2). Our data indicate that this widespread shift to dominant nannofossil oozes²⁰ was triggered by the motion of Adria out of the near-equatorial upwelling belt into less trophic tropical waters dominated by calcareous nannoplankton.

Palaeomagnetic evidence has previously been presented¹¹ for the swing to low latitudes of Adria by the latest Jurassic. We observe that this movement started earlier, in the Middle Jurassic, and was more pronounced than generally realized because, to our knowledge, palaeomagnetic data for Adria's radiolarites that were deposited at the lowest palaeolatitudes had not been reported before ref. 10 and this study. This interval of rapid plate motion is largely underestimated in conventional apparent polar wander paths (APWPs) based on running averages of palaeopole entries. For example, the most recent Besse and Courtillot²⁶ master APWP transferred to Africa in the interval 170–145 Myr ago includes palaeopoles from limestones from central Europe that have an ambiguous smeared distribution with a suspicion of remagnetization¹⁸, whereas well-dated Adria poles were excluded from the compilation even though the tectonic affinity of parts of Adria to Africa is well demonstrated^{11,12}. As a result, latitudes predicted for Colle Sogno by this master APWP show little variation over the Jurassic–Early Cretaceous (Fig. 2).

Modern climate zonality provides a first-order hypothesis to explain ancient sedimentary facies that can be tested with palaeomagnetism. Paleolatitudinal data from a continuous succession of sediments at Colle Sogno show that carbonate facies dominated when Adria was located at tropical latitudes in the Early Jurassic, radiolarian oozes were deposited when it moved to near-equatorial latitudes in the Middle–Late Jurassic, and nannofossil oozes became progressively dominant when it returned to higher tropical latitudes in the Late Jurassic–Cretaceous. This tectonic model for the genesis of radiolarites on Adria's rifted margins adds an essential horizontal plate motion component to explanations involving only vertical variations of the CCD relative to the ocean floor¹, which is an efficient mechanism for varying carbonate concentration rather than form radiolarian oozes. □

Methods

Samples and measurements

Standard high-field thermomagnetic experiments show that samples from the lower part of the succession (dominantly grey Zu Limestone to lower Sogno Formation) contain magnetite with moderate coercivity and maximum unblocking temperatures ($T_{\max}$) of ~570 °C in association with subsidiary haematite with higher coercivity and a $T_{\max}$ of ~680 °C (Supplementary Fig. A). A magnetite–haematite mixture dominates the grey to pink upper Sogno Formation, magnetite associated with subsidiary haematite characterizes the green and red radiolarites, a magnetite–haematite mixture dominates the reddish Rosso ad Aptici Formation, and magnetite alone characterizes the white Maiolica Formation.

All samples were thermally demagnetized and analysed at the palaeomagnetic laboratory of the University of Munich (Supplementary Table A). The extremely weak magnetization of the Colle Sogno samples (0.1–15 mA m⁻¹) required the use of highly sensitive magnetometers with d.c. SQUID technology. Demagnetization showed the initial removal of an overprint component usually aligned along the present-day magnetic field and a final characteristic component over higher demagnetization temperatures that is compatible with magnetite and occasionally haematite as the remanence carriers (Supplementary Fig. B). The characteristic magnetization, calculated by means of principal component analysis, has dual polarity, being oriented northerly-and-down or southerly-and-up (Supplementary Table B). It was observed in samples from all sites (except site CS7B) and showed no discernible change in orientation with respect to the origin of the axes of orthogonal projection plots in mixed mineralogy samples. Other magnetization components observed in some samples over intermediate unblocking temperatures (200–550 °C) are usually scattered between the present-day field and the characteristic magnetization directions, suggesting that they represent composite magnetizations, although overprinting during the Alpine orogeny may be a contributing factor.

Data analysis

The characteristic components of magnetization, grouped in three distinct stratigraphic intervals (Domaro Formation, radiolarites–basal Rosso ad Aptici, and upper Rosso ad Aptici–basal Maiolica) by means of standard Fisher statistics, pass the polarity reversal test (class C of ref. 27). The generally homogeneous bedding attitudes preclude the application of a fold test at Colle Sogno. Two magnetostratigraphically correlated sections from the Maiolica Formation at Alpetto and Piè del Dosso^{7,8}, located respectively west and east of Colle Sogno (Fig. 1), have characteristic mean directions that differ by 105° in geographic coordinates but only 4° in tilt-corrected coordinates. These tilt-corrected magnetizations, similar in orientation to the ones from the basal Maiolica Formation at site CS11, pre-date Alpine folding and suggest internal structural consistency of this part of the South Alpine belt. We therefore interpret the dual-polarity directions from Colle Sogno as primary magnetizations acquired in normal and reverse polarity geomagnetic fields.

Fisher statistics were applied to calculate site mean characteristic component directions. From the characteristic component of each sample, a virtual geomagnetic pole was calculated and Fisher statistics were applied to calculate site mean palaeomagnetic poles. Palaeolatitudes were calculated as 90° minus colatitude, where colatitude is the arc distance between the Colle Sogno average palaeomagnetic site and the individual palaeomagnetic north poles, with palaeolatitude error defined by $\pm A95$ (Fisher's cone of 95% confidence about the palaeomagnetic pole) (Supplementary Table C).

Fisher statistics were applied to calculate selected overall mean palaeomagnetic poles for Adria (Supplementary Table C) for palaeogeographic reconstructions (Fig. 2, right panels). Rotation poles of Gondwanan elements relative to northwest Africa (Adria) are from ref. 28, rotation poles of Laurasian elements relative to Adria are from ref. 29, and rotation poles of North America relative to Africa are from ref. 29 for pre-oceanic spreading time and from ref. 30 for syn-oceanic spreading times. The reconstructions at 166–149 Myr ago, 149–140 Myr ago and 127–119 Myr ago use, respectively, magnetic anomalies M25 (~155 Myr ago), M21–M16 (interpolated mid-value at ~144.5 Myr ago) and M0 (~125 Myr ago).

Received 17 June 2004; accepted 6 January 2005; doi:10.1038/nature03378.

- Winterer, E. L. & Bosellini, A. Subsidence and sedimentation on Jurassic passive continental margin, Southern Alps, Italy. *Bull. Am. Assoc. Petrol. Geol.* **65**, 394–421 (1981).
- Kennett, J. P. *Marine Geology* (Prentice-Hall, Englewood Cliffs, 1982).
- Baumgartner, P. et al. New Middle and Upper Jurassic radiolarian assemblages co-occurring with ammonites and nannofossils from Southern Alps (northern Italy). *Mem. Geol. (Lausanne)* **23**, 737–750 (1995).
- Erba, E. Nannofossils and Mesozoic oceanic anoxic events. *Mar. Micropaleontol.* **52**, 85–106 (2004).
- Gradstein, F. M. et al. A new geologic time scale, with special reference to Precambrian and Neogene. *Episodes* **27**, 83–100 (2004).
- Hornor, F. & Heller, F. Lower Jurassic magnetostratigraphy at the Breggia Gorge (Ticino, Switzerland) and Alpe Turati (Como, Italy). *Geophys. J. R. Astron. Soc.* **73**, 705–718 (1983).
- Channell, J. E. T., Cecca, F. & Erba, E. Correlations of Hauterivian and Barremian (Early Cretaceous) stage boundaries to polarity chrons. *Earth Planet. Sci. Lett.* **134**, 125–140 (1995).
- Channell, J. E. T. & Erba, E. Early Cretaceous polarity chrons CM0 to CM11 recorded in northern Italian land sections near Brescia. *Earth Planet. Sci. Lett.* **108**, 161–179 (1992).
- Channell, J. E. T., Erba, E. & Lini, A. Magnetostratigraphic calibration of the Late Valanginian carbon isotope event in pelagic limestones from Northern Italy and Switzerland. *Earth Planet. Sci. Lett.* **118**, 145–166 (1993).
- Aiello, I. W. & Hagstrum, J. T. Paleomagnetism and paleogeography of Jurassic radiolarian cherts from the northern Apennines of Italy. *Geol. Soc. Am. Bull.* **113**, 469–481 (2001).
- Channell, J. E. T. in *Paleomagnetism and Tectonics of the Mediterranean Region* (eds Morris, A. & Tarling, D. H.) 119–132 (Geological Society Special Publication, London, 1996).
- Muttoni, G. et al. Early Permian Pangea 'B' to Late Permian Pangea 'A'. *Earth Planet. Sci. Lett.* **215**, 379–394 (2003).
- Hargraves, R. B., Briden, J. C. & Daniels, B. A. Palaeomagnetism and magnetic fabric in the Freetown Complex, Sierra Leone. *Geophys. J. Int.* **136**, 705–713 (1999).
- Sichler, J. L. et al. Mobility of Morocco. *Can. J. Earth Sci.* **17**, 1546–1558 (1980).
- Hargraves, R. B. Paleomagnetism of Mesozoic kimberlites in Southern Africa and the Cretaceous apparent polar wander curve for Africa. *J. Geophys. Res.* **94**, 1851–1866 (1989).
- Van Fossen, M. C. & Kent, D. V. A palaeomagnetic study of 143 Ma kimberlite dikes in central New York State. *Geophys. J. Int.* **113**, 175–185 (1993).
- Van Fossen, M. C. & Kent, D. V. Paleomagnetism of 122 Ma plutons in New England and the Mid-Cretaceous paleomagnetic field in North America: true polar wander or large-scale differential mantle motion? *J. Geophys. Res.* **97**, 19651–19661 (1992).

18. Van der Voo, R. *Paleomagnetism of the Atlantic, Tethys and Iapetus Oceans* (Cambridge University Press, Cambridge, 1993).
19. De Wever, P., Azéma, J. & Fourcade, E. Radiolaires et radiolarites: production primaire, diagenèse et paléogéographie. *Bull. Centres Rech. Explor. Product. Elf Aquitaine* **18**, 315–379 (1994).
20. Baumgartner, P. O. Age and genesis of Tethyan Jurassic Radiolarites. *Eclog. Geol. Helv.* **80**, 831–879 (1987).
21. Bartolini, A., Baumgartner, P. O. & Guex, J. Middle and Late Jurassic radiolarian palaeoecology versus carbon-isotope stratigraphy. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **145**, 43–60 (1999).
22. Bernoulli, D., Manatschal, G., Desmurs, L. & Muentener, O. Where did Gustav Steinmann see the trinity? Back to the roots of an Alpine ophiolite concept. *Geol. Soc. Am. Spec. Pap.* **373**, 93–110 (2003).
23. Richards, D., Butler, R. F. & Harms, T. A. Paleomagnetism of the late Paleozoic Slide Mountain terrane, northern and central British Columbia. *Can. J. Earth Sci.* **30**, 1898–1913 (1993).
24. Hagstrum, J. T. & Murchev, B. L. Deposition of Franciscan Complex cherts along the paleoequator and accretion to the American margin at tropical paleolatitudes. *Geol. Soc. Am. Bull.* **105**, 766–778 (1993).
25. Larson, R. L., Steiner, M. B., Erba, E. & Lancelot, Y. Paleolatitudes and tectonic reconstructions of the oldest portion of the Pacific Plate: A comparative study. *Proc. ODP. Sci. Res.* **129**, 615–631 (1992).
26. Besse, J. & Courtillot, V. Correction to “Apparent and true polar wander and the geometry of the geomagnetic field over the last 200 Myr”. *J. Geophys. Res. Solid Earth* **108**, 2469, doi:10.1029/2003JB002684 (2003).
27. McFadden, P. L. & Lowes, F. J. The discrimination of mean directions drawn from Fisher distributions. *Geophys. J. R. Astron. Soc.* **67**, 19–33 (1981).
28. Lottes, A. L. & Rowley, D. B. in *Paleozoic Palaeogeography and Biogeography* (eds McKerrow, W. S. & Scotese, C. R.) 383–395 (Memoir 12, Geological Society of London, 1990).
29. Bullard, E. C., Everett, J. & Smith, A. G. The fit of the continents around the Atlantic. *Phil. Trans. R. Soc. Lond. A* **258**, 41–45 (1965).
30. Klitgord, K. D. & Schouten, H. in *The Geology of North America, The Western North Atlantic Region* (eds Vogt, P. R. & Tucholke, B. E.) 351–377 (Geological Society of America, Boulder, 1986).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements E. Dallanave is thanked for assistance in the field and laboratory analysis. Comments by D. Bernoulli and J. Hagstrum greatly improved the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.M. (giovanni.muttoni@unimi.it).

Insolation-driven changes in atmospheric circulation over the past 116,000 years in subtropical Brazil

Francisco W. Cruz Jr^{1,2}, Stephen J. Burns¹, Ivo Karmann², Warren D. Sharp³, Mathias Vuille¹, Andrea O. Cardoso⁴, José A. Ferrari⁵, Pedro L. Silva Dias⁴ & Oduvaldo Viana Jr²

¹Department of Geosciences, University of Massachusetts, Amherst, 01002 Massachusetts, USA

²Instituto de Geociências, Universidade de São Paulo, Rua do Lago 562, São Paulo, SP, 05508-080, Brazil

³Berkeley Geochronology Center, 2455 Ridge Rd, Berkeley, California 94709, USA

⁴Departamento de Ciências Atmosféricas, Instituto de Astronomia, Geofísica e Ciências Atmosféricas, Universidade de São Paulo, Rua do Matão 1226, São Paulo, SP 05508-090, Brazil

⁵Instituto Geológico, Av. Miguel Stefano 3900, São Paulo, SP 04301-903, Brazil

During the last glacial period, large millennial-scale temperature oscillations—the ‘Dansgaard/Oeschger’ cycles—were the primary climate signal in Northern Hemisphere climate archives from the high latitudes to the tropics^{1–6}. But whether the influence of these abrupt climate changes extended to the tropical and subtropical Southern Hemisphere, where changes in insolation are thought to be the main direct forcing of climate, has remained unclear. Here we present a high-resolution oxygen isotope record of a U/Th-dated stalagmite from subtropical southern Brazil, covering the past 116,200 years. The oxygen isotope signature varies with shifts in the source region and amount of rainfall in the area, and hence records changes in

atmospheric circulation and convective intensity over South America. We find that these variations in rainfall source and amount are primarily driven by summer solar radiation, which is controlled by the Earth’s precessional cycle. The Dansgaard/Oeschger cycles can be detected in our record and therefore we confirm that they also affect the tropical hydrological cycle, but that in southern subtropical Brazil, millennial-scale climate changes are not as dominant as they are in the Northern Hemisphere.

The study area in subtropics of southeastern Brazil is well-situated for investigating changes in tropical and subtropical atmospheric circulation. Although wet and dry seasons are not observed, the region receives rainfall from two distinct sources in different seasons (see Supplementary Information S1). During the austral winter and early spring, equatorward incursions of mid-latitude cold dry air result in cyclonic storms that move moisture inland from the nearby Atlantic Ocean⁷. During the late summer and early autumn, rainfall is related to the intense convection over the interior Amazon basin that is associated with the South American summer monsoon (SASM)⁸. From December to March, a northwesterly low-level flow, the Andean low-level jet (ALLJ), transports Amazon tropical moisture from interior Brazil near the Equator towards the South Atlantic Convergence Zone (SACZ) located over southern Brazil⁹. At the study site (27°S), a significant fraction of summer precipitation derives from this moisture flux¹⁰.

Thus, although it is not directly beneath the centre of convective activity over the Amazon basin, summer rainfall in southeastern Brazil is strongly influenced by the southward progression of convection across the Amazon basin through the summer. The two sources of rainfall also have quite distinct oxygen isotopic ratios. The more locally sourced winter rainfall is enriched in ¹⁸O compared to summer precipitation, with average $\delta^{18}\text{O}$ values of precipitation during the winter of −3‰ as compared to average values of −7‰ in the early autumn at Porto Alegre City, southern Brazil¹¹ (and see Supplementary Information S2). At present the mean annual isotopic composition of rainfall is mainly determined by the relative contribution of summer, monsoonal precipitation versus winter, extratropical precipitation and not by temperature or amount effects. This interpretation is supported by the coincidence of more positive values of $\delta^{18}\text{O}$ with the rainiest period in Porto Alegre during winter and early spring, which is contrary to the expected tendency if either temperature or precipitation amount were the major source of variation in rainfall oxygen isotope ratios¹².

Stalagmite BT2 was collected from Botuverá Cave (27° 13′ 24″ S; 49° 09′ 20″ W, 230 m above sea level, a.s.l.). On the basis of 20 U/Th analyses, the 70-cm-long stalagmite was deposited from ~116 thousand years (kyr) ago to the present without detectable hiatuses (Fig. 1). Samples for stable isotopic analyses were taken every 1 mm, which represents an average resolution of ~150 yr. Values of $\delta^{18}\text{O}$ for stalagmite BT2 range from −0.5 to −5.0 with an apparent cyclicity of ~20 kyr (Fig. 2). The lowest values are observed around 14–20 kyr ago and around 40–45 kyr ago. Superimposed on the longer-term cyclicity are more abrupt millennial-scale variations with an amplitude of ~1 to 1.5‰.

Stalagmite BT2 appears to have been deposited in approximate isotopic equilibrium with cave drip water as indicated by the relatively small ranges of $\delta^{18}\text{O}$ along single speleothem (such as stalactites or stalagmites) layers¹³ (S3a) and absence of significant correlations between $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ (Supplementary Information S3b and c). The relatively large range of variation in $\delta^{18}\text{O}$, more than 5‰ in the speleothem, also suggests that temperature, through its effect in the calcite–water fractionation factor, is not the primary cause of the observed variation. The temperature-dependent fractionation between calcite and water is −0.24‰ per °C (ref. 14), requiring 20 °C of temperature change to explain the total range of variation in $\delta^{18}\text{O}$, or more than twice what is thought to be the total

temperature change between glacial and interglacial climates in the subtropics^{15,16}. Furthermore, the colder temperatures of the glacial period at ~18 kyr ago would result in more positive calcite $\delta^{18}\text{O}$ values, not more negative as is observed during the Last Glacial Maximum. We conclude that $\delta^{18}\text{O}$ of speleothem BT2 primarily reflects changes in $\delta^{18}\text{O}$ of regional precipitation.

Figure 2 shows a comparison of the BT2 isotopic time series with incoming solar radiation for the month of February at 30° S. Variation in the latter is dominated by changes in precession of the Earth's orbit¹⁷. For each of the last five precessional cycles, minima and maxima in solar radiation align remarkably well with maxima and minima, respectively (the scale for insolation in Fig. 2 is reversed), in calcite $\delta^{18}\text{O}$. In addition, amplitude of the isotopic variation matches that of insolation. Spectral analysis of the BT2 record shows a dominant peak in spectral power at 23 kyr (see Supplementary Information S4), as does precession. By far the majority of the variance in $\delta^{18}\text{O}$ is related to changes in insolation. Insolation could affect the isotopic composition of precipitation, and stalagmite BT2, in two main ways.

First, on the basis of modern isotopic climatology, we interpret increases in the isotopic values to reflect a greater proportion of winter versus summer rainfall. This seasonal balance in precipitation is in turn controlled by the long-term mean location and southward extent of convective activity associated with the South American summer monsoon and southern boundary of the Hadley cell in the Southern Hemisphere. Because of its location, the isotopic record of stalagmite BT2 should also record changes in the location of the SACZ. Indeed, the SACZ is the exit region of the monsoon low-level winds and its intensity during summer reflects the moisture convergence associated with this monsoon flow¹⁸. During minima in summer solar radiation in the Southern Hemisphere tropics and subtropics, the mean location of the SASM and the SACZ may shift northward and less Amazon basin moisture will be captured and transported towards the southeast, decreasing the relative contribution of summer monsoonal rainfall. The opposite will be true when summer insolation is at a maximum.

Second, insolation could influence the isotopic composition of rainfall and stalagmite BT2 by affecting the intensity of convection

in both SASM and the SACZ. Convective intensity could affect the isotopic composition of rainfall through an amount effect. Speleothem $\delta^{18}\text{O}$ values, particularly those from the tropics or subtropics^{5,6}, are often interpreted as indicating changes in the amount of rainfall because rainfall amount and $\delta^{18}\text{O}$ are observed to be inversely correlated in regions of deep convection¹⁹. Periods of increased summer insolation in the region should be periods of increased convection, increased precipitation and decreased $\delta^{18}\text{O}$ values of that precipitation. In fact, today $\delta^{18}\text{O}$ over southern Brazil (Porto Alegre) is significantly correlated with monsoon intensity and the southward extent of the SASM²⁰. During periods of increased convection and precipitation associated with an intensified SASM, the remaining water vapour—which is subsequently transported towards the SACZ by the ALLJ—is more depleted in $\delta^{18}\text{O}$. An intensified SASM over tropical South America leads to enhanced tropical convection and preferential rainout of ^{18}O over the Amazon and ultimately to more depleted $\delta^{18}\text{O}$ downstream²⁰. Thus, changes in convective intensity within the SASM/SACZ may significantly contribute to and reinforce the effect that the location of this system has on the $\delta^{18}\text{O}$ of precipitation. The BT2 record probably reflects both of these processes and suggests that both effects are primarily steered by precession-driven changes in solar insolation. Maxima in solar insolation in Southern Hemisphere summer result in maximum movement of the SASM/SACZ to the south and intensified convective activity in this system.

A comparison of the Holocene portion of our record with other studies of climate change in South America supports this interpretation. The BT2 isotopic record becomes progressively more negative over the course of the Holocene, during which time palynological studies indicate expansion of the Atlantic rainforest in coastal regions of southern Brazil²¹ and a southward expansion of the Amazon rainforest along its southwestern border²². Studies of lakes on the Peruvian and Bolivian Altiplano also indicate increasing available moisture, sourced from the Amazon basin, over the course of the Holocene^{23,24}.

The $\delta^{18}\text{O}$ values of stalagmite BT2 also show a strong correspondence with atmospheric methane concentrations (Fig. 2) as determined from ice-core records in Greenland²⁵. Rapid increases in methane concentrations are coincident with rapid increases in $\delta^{18}\text{O}$

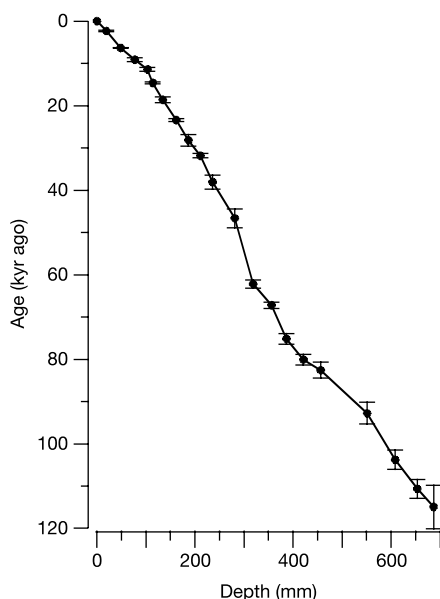


Figure 1 Age versus depth model for stalagmite BT2. Error bars are 95% confidence limits. The solid line connecting the data points is the linear interpolation used to calculate ages of individual oxygen isotope data points.

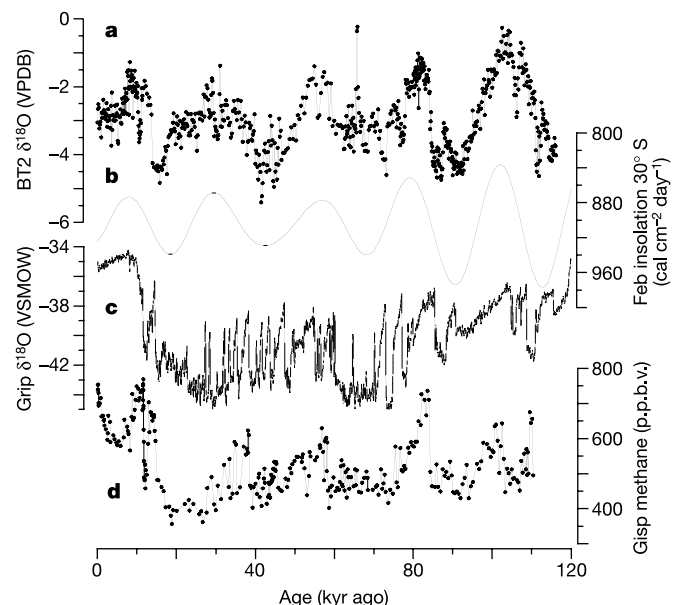


Figure 2 Stable oxygen isotope profile for stalagmite BT2. The BT2 profile (a) is compared with February solar insolation for 30° S (b), oxygen isotopes of the NGrp ice core from Greenland (c), and atmospheric methane concentrations from the Greenland ice core (d).

in our record and, if our interpretation is correct, with rapid northward shifts in the mean location of the South American monsoon. We suggest that during these periods the centre of convective activity related to the SASM was probably displaced northward, resulting in increased moisture and increased methane production in tropical wetlands. In addition to the movement of the monsoon a general intensification of the mean zonal Hadley circulation and tropical convective activity may have accompanied these shifts. This mechanism is also valid for the short-term negative trend in BT2 during the Younger Dryas period, which is consistent with rapid southward movement of the South American monsoon. During this time there is evidence of significant decrease in discharge of the Amazon River due to decreased monsoon precipitation in lowland tropical South America²⁶.

Although precessional-scale changes in solar radiation appear to be the dominant driver of climate change in the region, other factors must also be involved. First, changes in the $\delta^{18}\text{O}$ value of sea water also occurred and serve to dampen slightly the precession-driven changes. For example, seawater $\delta^{18}\text{O}$ decreased by $\sim 1\%$ during the transition from the Last Glacial Maximum to the Holocene²⁷, during which time the BT2 isotopic values increased by $\sim 3\%$. Second, boundary conditions, in particular Northern Hemisphere ice volume and temperature, probably have an additional effect on the mean location of the South American monsoon that is superimposed on precession-driven changes even in the Southern Hemisphere. For example, the isotopic values of BT2 at ~ 17 kyr ago and ~ 40 kyr ago are about 1.5% more negative than at present, yet each of these three periods are close to maxima in solar radiation for Southern Hemisphere summers. Modelling studies indicate that during times of increased ice volume and decreased temperatures in the high northern latitudes, the intertropical convergence zone is pushed farther south than would be dictated by precession alone²⁸. We suggest that the same is likely to be true for the SASM/SACZ system as a necessary response of the Hadley circulation in increasing its northward heat transport to compensate for the loss of energy caused by increased ice volume at high northern latitudes. In the study area, the result is an increase in the relative amount of monsoon-related precipitation in summer. Such a mechanism could explain, for example, the short, negative isotopic excursion in the BT2 record that occurs at the time of the Younger Dryas cold event in the Northern Hemisphere.

Figure 2 also compares the BT2 $\delta^{18}\text{O}$ values with isotopic data from the Grip Greenland ice core²⁹. In the ice core, as in a variety of climate archives from the Northern Hemisphere from this time period, isotopic variation during the last glacial period is dominated by Dansgaard/Oeschger (D/O) cycles. In the BT2 time series, D/O-type cycles are observable as rapid $1\text{--}2\%$ excursions in the isotopic values, although it is not possible to identify with certainty each individual D/O event in the BT2 record. The ages of more prominent, and more easily identifiable D/O events in the BT2 record match the absolute chronology of a speleothem record from Hulu Cave in China⁶ quite well. In contrast to Northern Hemisphere records, however, in which D/O events are the primary mode of climate variation, the importance of D/O events is greatly reduced in our Southern Hemisphere record and precession-controlled changes in solar insolation dominate long-term variability in the tropical hydrological cycle. □

Methods

Age determinations were carried out at Berkeley Geochronology Center, using conventional chemical and thermal ionization mass spectrometry (TIMS) techniques³⁰. Twenty samples, weighting between 409 and 440 mg were cleaned ultrasonically in alcohol, totally dissolved by attack with concentrated HNO_3 and equilibrated with a $^{236}\text{U}\text{--}^{233}\text{U}\text{--}^{229}\text{Th}$ spike. U and Th were separated by ion exchange columns, loaded onto outgassed rhenium filaments, and measured on a VG-Sector 54 mass spectrometer equipped with a high abundance-sensitivity filter and Daly-ion counter. Measured isotope ratios were corrected for minor amounts of initial U and Th using ^{232}Th as an index isotope and assuming a typical silicate composition for the contaminant; that is, activity

ratios of $^{232}\text{Th}/^{238}\text{U} = 1.21 \pm 0.6$, $^{230}\text{Th}/^{238}\text{U} = 1.0 \pm 0.1$, and $^{234}\text{U}/^{238}\text{U} = 1.0 \pm 0.1$. U–Th isotopic data and ages are shown in Supplementary Table 1. Age errors are 95% confidence limits.

Oxygen isotope ratios are expressed in δ notation, the per mil deviation from the Vienna Pee-Dee Belemnite standard. For example, for oxygen, $\delta^{18}\text{O} = [((^{18}\text{O}/^{16}\text{O})_{\text{sample}}/(^{18}\text{O}/^{16}\text{O})_{\text{VPDB}}) - 1] \times 1,000$. For each measurement, approximately 200 μg of powder was drilled from the sample and analysed with an on-line, automated, carbonate preparation system linked to a Finnigan Delta XL ratio mass spectrometer at the University of Massachusetts. Reproducibility of standard materials is 0.08% for $\delta^{18}\text{O}$.

Received 16 August 2004; accepted 11 January 2005; doi:10.1038/nature03365.

- Dansgaard, W. *et al.* Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* **364**, 218–220 (1993).
- Peterson, L. C. *et al.* Rapid changes in the hydrologic cycle of the tropical Atlantic during the Last Glacial. *Science* **290**, 1947–1951 (2000).
- Schulz, H., von Rad, U. & Erlenkeuser, H. Correlation between Arabian Sea and Greenland climate oscillations of the past 110,000 years. *Nature* **393**, 54–57 (1998).
- Allen, J. R. M. *et al.* Rapid environmental changes in southern Europe during the last glacial period. *Nature* **400**, 740–743 (1999).
- Burns, S. J., Fleitmann, D., Kramers, J., Matter, A. & Al-Subbar, A. Indian Ocean climate during the last glacial period and an absolute chronology for Dansgaard/Oeschger oscillations 9 to 13. *Science* **301**, 1365–1367 (2003).
- Wang, Y. J. *et al.* A high-resolution absolute-dated late Pleistocene monsoon record from Hulu Cave, China. *Science* **294**, 2345–2348 (2001).
- Vera, C. S., Vigliarolo, P. K. & Berbery, E. H. Cold season synoptic-scale waves over subtropical South America. *Mon. Weath. Rev.* **130**, 684–699 (2002).
- Zhou, J. & Lau, K. M. Does a monsoon climate exist over South America? *J. Clim.* **11**, 1020–1040 (1998).
- Gan, M. A., Kousky, V. E. & Ropelewski, C. F. The South American monsoon circulation and its relationship to rainfall over West-Central Brazil. *J. Clim.* **17**, 47–66 (2004).
- Rao, V. B., Cavalcanti, I. F. A. & Hada, K. Annual variation of rainfall over Brazil and water vapor characteristics over South America. *J. Geophys. Res.* **101**, 26539–26551 (1996).
- IAEA/WMO. *Global Network for Isotopes in Precipitation (GNIP) Database* (IGBP PAGES/World Data Center-A for Paleoclimatology Data Contribution Series 94–005 NOAA/NGDC Paleoclimatology Program, Boulder, Colorado, 1994); (<http://isohis.iaea.org>).
- Rozanski, K., Araguás-Araguás, L. L. & Gonfiantini, R. In *Climate Change in Continental Isotopic Records* (eds Swart, P. K., Lohmann, K. C., McKenzie, J. & Savin, S.) 1–36 (Geophysical Monograph 78, American Geophysical Union, Washington, DC, 1993).
- Hendy, C. H. The isotopic geochemistry of speleothems. I. The calculation of the effects of different modes of formation on the isotopic composition of speleothems and their applicability as paleoclimatic indicators. *Geochim. Cosmochim. Acta* **35**, 801–824 (1971).
- Friedman, I. & O'Neil, J. R. In *Data of Geochemistry* (ed. Fleischer, E. M.) Ch. KK, 1–12 (United States Geological Survey Professional Paper 440-KK, US Government Printing Office, Washington, DC, 1977).
- Stute, M. *et al.* Cooling of tropical Brazil (5°C) during the last glacial maximum. *Science* **269**, 379–383 (1995).
- Lea, D., Pak, D. K., Peterson, L. C. & Hughes, K. A. Synchronicity of tropical and high latitude temperature over the last glacial termination. *Science* **301**, 1361–1364 (2003).
- Berger, A. & Loutre, M. F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10**, 297–317 (1991).
- Nogués-Paegle, J. & Mo, K. C. Alternating wet and dry conditions over South America during summer. *Mon. Weath. Rev.* **125**, 279–291 (1997).
- Gat, J. R. Oxygen and hydrogen isotopes in the hydrologic cycle. *Annu. Rev. Earth Planet. Sci.* **24**, 225–262 (1996).
- Vuille, M., Bradley, R. S., Werner, M., Healy, R. & Keimig, F. Modeling $\delta^{18}\text{O}$ in precipitation over the tropical Americas: 1. Interannual variability and climatic controls. *J. Geophys. Res.* **108**, 4174, doi:10.1029/2001JD002038 (2003).
- Behling, H. South and Southeast Brazilian grasslands during Late Quaternary times: a synthesis. *Palaeogeogr. Palaeoclim. Palaeoecol.* **177**, 19–27 (2002).
- Mayle, F. E., Burbridge, R. & Killeen, T. J. Millennial-scale dynamics of southern Amazonian rain forests. *Science* **290**, 2291–2294 (2000).
- Baker, P. A. *et al.* The history of South American tropical precipitation for the past 25,000 years. *Science* **291**, 640–643 (2001).
- Seltzer, G., Rodbell, D. & Burns, S. J. Isotopic evidence for Late Glacial and Holocene hydrologic change in tropical South America. *Geology* **28**, 35–38 (2000).
- Brook, E. J., Sowers, T. & Orchard, J. Rapid variations in atmospheric methane concentration during the past 110,000 years. *Science* **273**, 1087–1091 (1996).
- Maslin, M. A. & Burns, S. J. Reconstruction of the Amazon Basin effective moisture availability over the past 14,000 years. *Science* **290**, 2285–2287 (2000).
- Schrag, D. *et al.* The oxygen isotopic composition of seawater during the Last Glacial Maximum. *Quat. Sci. Rev.* **21**, 331–342 (2002).
- Chiang, J. C. H., Biasutti, M. & Battisti, D. S. Sensitivity of the Atlantic Intertropical Convergence Zone to the Last Glacial Maximum boundary conditions. *Paleoceanography* **18**(4), 1094, doi: 10.1029/2003PA000916 (2003).
- North Greenland Ice Core Project members. High resolution record of Northern Hemisphere climate extending into the last interglacial period. *Nature* **431**, 147–151 (2004).
- Sharp, W. D., Ludwig, K. R., Chadwick, O. A., Amundson, R. & Glaser, L. L. Dating fluvial terraces by $^{230}\text{Th}/\text{U}$ on pedogenic carbonate, Wind River Basin, Wyoming. *Quat. Res.* **59**, 139–150 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) for financial support, the CECAV/IBAMA for authorizing cave samples collection and the GEEP-Açungui Speleological Group and E. Barni for supporting field work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.W.C. Jr (fdacruz@geo.umass.edu).

Water-rich basalts at mid-ocean-ridge cold spots

Marco Ligi¹, Enrico Bonatti^{1,2,3}, Anna Cipriani^{1,2} & Luisa Ottolini⁴

¹Istituto Scienze Marine, Geologia Marina, CNR, Via Gobetti 101, 40129 Bologna, Italy

²Lamont Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

³Dipartimento di Scienze della Terra, Università "La Sapienza", Piazzale Aldo Moro 5, 00187 Rome, Italy

⁴Istituto di Geoscienze e Georisorse, Sezione di Pavia, CNR, Via Ferrata 1, 27100 Pavia, Italy

Although water is only present in trace amounts in the sub-oceanic upper mantle, it is thought to play a significant role in affecting mantle viscosity, melting and the generation of crust at mid-ocean ridges. The concentration of water in oceanic basalts^{1,2} has been observed to stay below 0.2 wt%, except for water-rich basalts sampled near hotspots and generated by 'wet' mantle plumes^{3–5}. Here, however, we report unusually high water content in basaltic glasses from a cold region of the mid-ocean-ridge system in the equatorial Atlantic Ocean. These basalts are sodium-rich, having been generated by low degrees of melting of the mantle, and contain unusually high ratios of light versus heavy rare-earth elements, implying the presence of garnet in the melting region. We infer that water-rich basalts from such regions of thermal minima derive from low degrees of 'wet' melting greater than 60 km deep in the mantle, with minor

dilution by melts produced by shallower 'dry' melting—a view supported by numerical modelling. We therefore conclude that oceanic basalts are water-rich not only near hotspots, but also at 'cold spots'.

The water content of the oceanic upper mantle can be estimated from the water concentration in mid-ocean-ridge basalt (MORB) glasses, after correcting for the effects of degassing and magmatic differentiation. The H₂O content of normal MORB (N-MORB) is generally below 0.2 wt% (ref. 1). Given that H₂O is about as incompatible as Ce, and assuming ~10% average degree of melting of the mantle upwelling below mid-ocean ridges (MORs), the mantle source of N-MORB is assumed to contain 0.01–0.02 wt% H₂O (ref. 2). However, basalts from topographically swollen portions of MORs have H₂O concentrations higher than those of N-MORB (Fig. 1). These swollen ridges are generally interpreted as being influenced by hot plumes rising from the transition zone or from even deeper in the mantle. Thus, the H₂O content of the mantle source of plume-type oceanic basalts is probably significantly higher than that of the N-MORB source region. For example, the mantle source of the Icelandic³ and Azores platform⁴ crust contains between 620 and 920 p.p.m. H₂O, that is, several times higher than that of the N-MORB source. Concerning off-ridge hotspots, an H₂O content of 405 ± 190 p.p.m. has been estimated for the mantle source of Hawaiian basalts⁵, supporting the hypothesis that plume-type mantle is H₂O-rich relative to the N-MORB mantle source. High water and volatile contents lower the mantle solidus, so that the mantle melts deeper and to a higher degree during its ascent below MORs.

We report here that the H₂O content of basaltic glasses from the equatorial Mid-Atlantic Ridge (MAR) is significantly higher than that of N-MORB. However, these H₂O-rich basalts are associated not with a 'hot' portion of MOR, but with the opposite—that is, a thermal minimum in the ridge system. We will discuss a model that

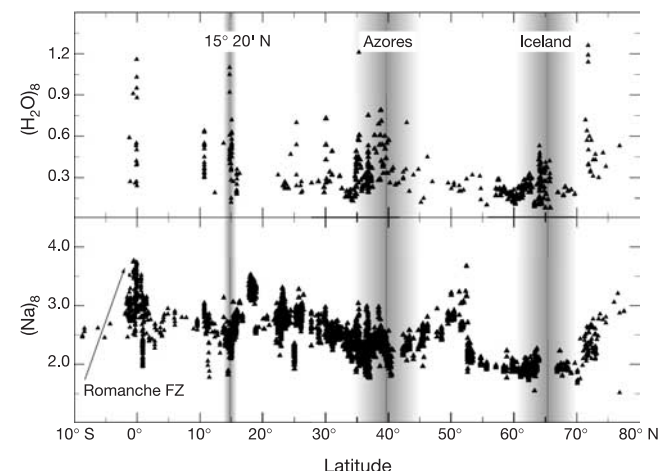


Figure 1 Distribution of Na₂O and H₂O in MORB glasses along the axis of the Mid-Atlantic Ridge from Iceland to the Equator. Data are from the Supplementary Table, our unpublished results and the Petrological Database of the Ocean Floor (PETDB) of Lamont Doherty Earth Observatory. FZ, fracture zone.

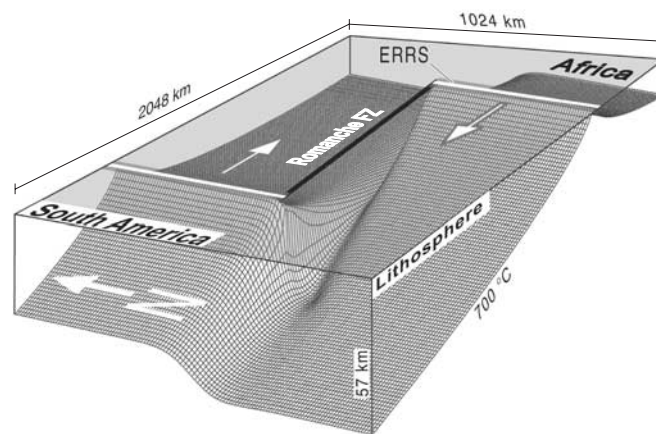


Figure 2 Geometry of the passive-flow model. The bases of the rigid plates represent the upper boundary layer in our plate-thickening passive mantle flow model. The model shown was obtained by iteratively solving the mantle temperature field at each time step, starting from a constant-thickness plate-flow model. The computed thickness of the African plate lithosphere at the ERRS transform intersection is ~50 km. Mantle flow velocities were estimated by solving a steady-state corner flow in a computational frame 2,048 × 1,024 km wide and 150 km deep (using a grid with 2 × 2 km spacing for each 1 km depth increment), assuming an incompressible homogeneous and isoviscous mantle. Mantle temperatures were predicted by solving the steady-state advection-diffusion equation, using an over-relaxation upwind finite difference method with variable grid spacing (512 × 256 × 101), and highest grid resolution (1 km) at the plate boundaries²⁸. Temperature solutions were found assuming 0 °C at the sea floor and 1,330 °C at 150 km depth, assuming the presence of an equatorial MAR cold spot^{11,12}.

explains why the H_2O content of oceanic basalt is high not only at hotspots, but also at 'cold spots'.

Basaltic glasses were sampled by dredging at several sites along the ~220-km-long MAR segment (the eastern Romanche ridge segment, ERRS) that extends south of the 900-km Romanche transform (Figs 2 and 3). Glasses were selected for freshness, and analysed by ion probe for rare-earth elements (REEs) and H_2O , and by electron microprobe for major elements (see Methods). The H_2O content of our glasses ranges from 0.25 to 1.10 wt% (see Supplementary Table). Their depth of eruption is >2,100 m below sea level; thus, they must be undersaturated in H_2O at these depths⁶ with little or no H_2O loss during eruption. The absence of vesicles in the glasses supports this conclusion. Their Cl content is <0.11% and mostly below the limit of detection (0.04%), suggesting no contamination by sea water. In order to correct for the effects of differentiation, we calculated $(\text{H}_2\text{O})_8$ —that is, H_2O normalized to 8 wt% MgO (refs 7, 8)—assuming olivine-plagioclase-clinopyroxene fractionation. The correction lowers somewhat the H_2O values but does not affect relative trends (see Supplementary Table).

$(\text{H}_2\text{O})_8$ and Na_8 plotted versus latitude along the MAR axis (Fig. 1) reveal variations of basalt water content, with maxima in regions where the MAR is affected by mantle plumes, such as at 62°–70° N (Iceland), 35°–45° N (Azores) and 15° 20' N. Maxima in H_2O content are generally mirrored by minima in Na_8 (Fig. 1), consistent with the idea whereby plume-related high degree of melting and water-rich plume mantle source go together^{9,10}. Glasses from the equatorial MAR are an exception to this pattern, in so far as they are H_2O -rich while Na_8 is also high. High- Na_8 basalts are consistent with a low degree of melting in this region^{11,12}. Peridotite mineral composition also suggests that the mantle in the equatorial MAR underwent exceptionally low (<5%) degrees of melting¹³, probably owing to the combined effects of a regional equatorial Atlantic thermal minimum^{11,12}, and a strong 'transform cold-edge effect'¹⁴, which cools the ridge as it approaches old/thick/cold lithosphere across transform offsets.

We carried out numerical experiments to estimate the extent to which the upper mantle is cooled by a long-offset, low-slip transform, such as the Romanche, and how partial melting and H_2O distribution are affected. We assumed a 900-km-long transform

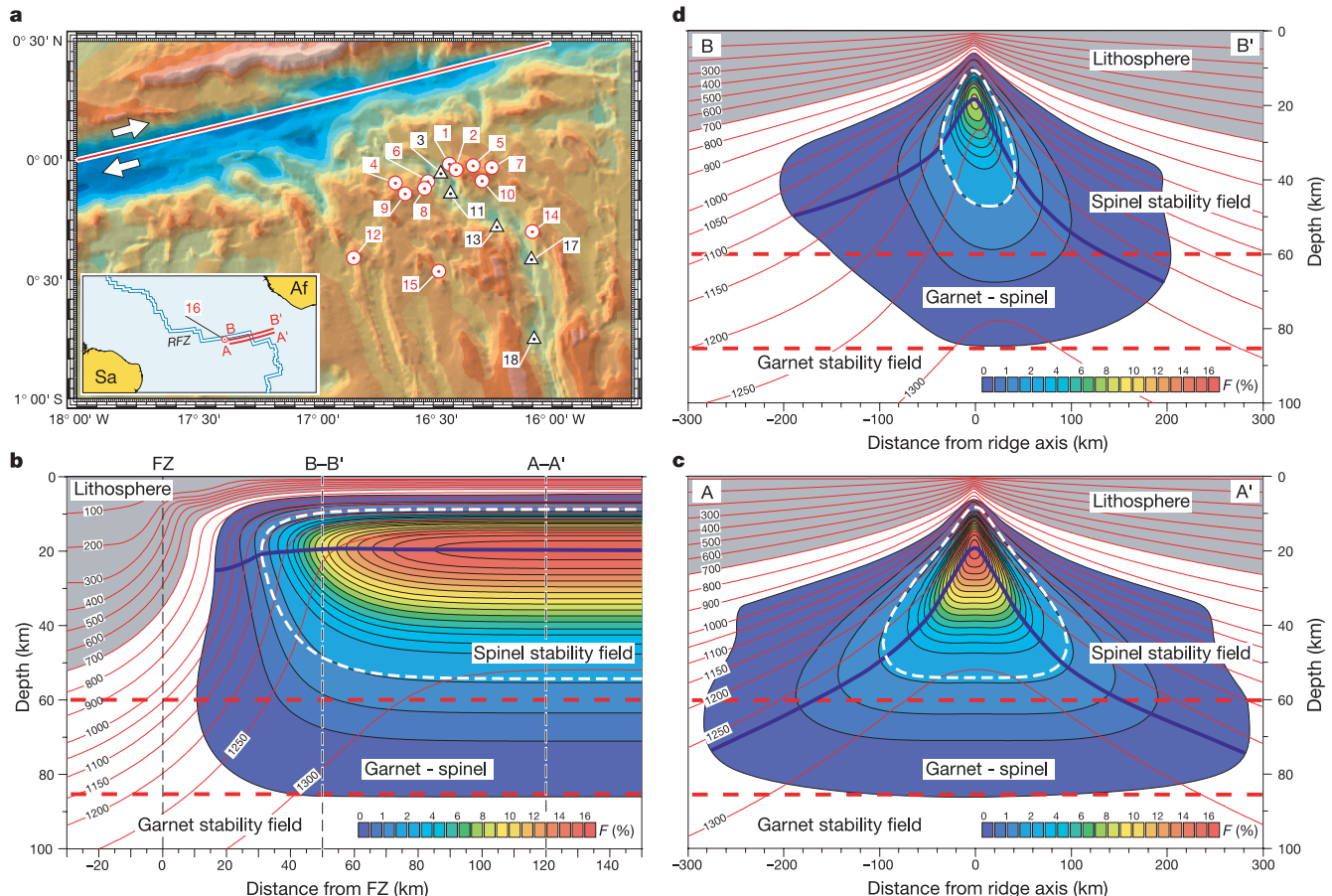


Figure 3 Multibeam topography of the eastern Romanche ridge–transform intersection and predicted melt production beneath the ERRS. **a**, Shaded relief image based on multibeam data. Depth ranges from 7,800 m (dark blue) to 1,000 m (light grey). Spreading direction and small circle path (thick red solid line) have been computed using the Africa–South America eulerian vector of the NUVEL-1A model²⁹. Sample locations (Supplementary Table) are indicated by red circles (this work) and black triangles. Numbers refer to position of samples in the Supplementary Table. Inset: RFZ, Romanche fracture zone; A–A' and B–B', location of sections shown in **c** and **d**. **b**, Fraction of melt (F , colour coded) generated along the ERRS axis, including the effect of water on peridotite solidus. The shape of the melting region is affected by mantle flow regimes (passive or

active). We neglected the effects of viscosity heterogeneity and buoyancy in our numerical experiments. The thick white dashed line marks the region of dry melting. The thick dark purple solid line marks the upper boundary of the region of melt production, that is, where the production rate is positive. Thick red dashed lines indicate boundaries between garnet and spinel stability fields. Mineral proportions in the transition zone between 85 and 60 km are assumed to vary linearly from pure garnet peridotite to pure spinel peridotite. Isotherms (°C) are indicated by thin red lines. **c**, **d**, Across-axis sections showing fraction of melt generated at the ERRS centre (120 km; **c**) and in the proximity (50 km) of the ridge–transform intersection (**d**).

offset, a half spreading rate of 16 mm yr^{-1} and that the bases of rigid plates were determined by the 700°C isotherm (Fig. 2). The ridge segment intersecting the transform was assumed to be 512 km long, longer than the real ERRS, in order to evaluate how far the transform effect extends along axis, avoiding numerical edge effects. Assuming a H_2O content of $\sim 175 \text{ p.p.m.}$ in the upper mantle¹⁵ the peridotite solidus is lowered, causing partial melting in a subridge mantle region that is wider and deeper than it would be if the mantle were dry^{16,17}. We modelled melt generation, including the effect of H_2O on the peridotite solidus, using a modified parameterization of experimental data from ref. 18. Batch and near-fractional melting are assumed, and simulated by mapping the melting interval from the batch melting experiments; water is treated as an incompatible component with a bulk distribution coefficient $D_{\text{H}_2\text{O}}$ that varies with pressure¹⁵ (see Supplementary Discussion).

Our numerical results (Fig. 3) show a strong decrease of 'crustal production' as the ridge approaches the transform, and no melting at all in a 20–40-km-wide strip close to the fracture zone (Fig. 3b), in agreement with the observation that the basaltic crust is nearly absent in that strip¹³. A cross-ridge subtriangular melting region, $\sim 600 \text{ km}$ wide at its base, is predicted beneath the centre of the segment, where the maximum degree of melting is 16.5% (Fig. 3c).

The melting region becomes smaller and asymmetric moving towards the ridge–transform intersection, with the maximum degree of melting decreasing rapidly (to $\sim 8\%$ at 50 km from the ridge–transform intersection), and the initial depth of melting varying greatly across axis (Fig. 3d).

Water addition deepens the onset of melting to 85 km beneath the centre of the segment, too cold for the anhydrous solidus to encounter garnet peridotite. Thus, water addition allows a significant melt fraction to be generated in the presence of residual garnet (Fig. 3c). In the proximity of the ridge–transform intersection, the partially molten region is mostly due to hydrous melting. The numerical model predicts that basalts sampled close to the Romanche fracture zone are generated exclusively in the subridge 'wet melting' mantle interval, that is, between ~ 80 and 60 km depth, within the region of stability of garnet. We would thus expect a significant 'garnet signature' in the chemical composition of our basalts, because they are undiluted by melts produced in the 'dry melting interval', above $\sim 60 \text{ km}$ depth within the spinel stability field.

REE partition coefficients during melting are different across the 80–60-km-deep boundary between garnet stability below, and spinel stability above: the heavy REEs are compatible with garnet but not with spinel. Thus, melting in the garnet stability field produces liquids depleted of heavy REEs relative to light REEs, and with chondrite-normalized Sm/Yb ratios, $(\text{Sm/Yb})_n$, of well over one^{19–21} and increasing with the proportion of melt generated in the garnet stability field. The deeper the level where the ascending mantle stops melting, the higher the proportion of melt generated in the presence of garnet²².

The concentration of highly incompatible elements in the aggregated liquid should be inversely proportional to the mean degree of melting. Therefore, incompatible elements should be increasingly enriched moving along axis towards the ridge–transform intersection. We calculated crustal thickness, mean pressure of melting, mean degree of melting, and mean composition of the aggregate melt, at any location along axis from the centre towards the tip of the ERRS, for each of the following melting models: wet and dry; batch; near-fractional; and pure-fractional. Mantle mineral assemblages for garnet, spinel and plagioclase peridotite are those of ref. 23; REE source composition and distribution coefficients are from ref. 24.

Basalt Na_8 , $(\text{Sm/Yb})_n$ and H_2O contents increase along axis towards the ridge–transform intersection, as predicted by the numerical model (Fig. 4). Note that when hydrous melting is included, the selected melting regime (batch, pure- or near-fractional) affects melting parameter predictions, owing to the pressure release melt parameterization adopted. The observed along-axis average patterns of melting parameter chemical indicators, such as Na_8 , Fe_8 and REE concentrations (Fig. 4), suggest that a pure-fractional or near-fractional melting model with a very low residual porosity ($<0.5\%$) fits the data best.

We conclude that the H_2O content of the oceanic basaltic crust peaks not only close to hotspots, but also at 'cold spots' along MORs. However, whereas hotspot H_2O maxima are caused by high degrees of melting of their H_2O -rich mantle plume sources, the H_2O enrichment of 'cold spots' is due to low degrees of melting occurring mostly within the 'wet melting' depth interval below the ridge, largely within the garnet stability field, with minor dilution from shallower 'dry' melts. Our results are consistent with the ubiquitous presence in the deeper part of the subridge melting column of volatiles and enriched components that are tapped preferentially during incipient 'wet' melting, but which are normally diluted by more abundant 'dry' melts generated in the shallower part of the melting column.

Extrapolating from these results, we expect relatively high H_2O content in basaltic crust generated at other 'thermal minima' along the MOR system, as at the Australian/Antarctic discordance²⁵ and at the Gakkel ridge²⁶.

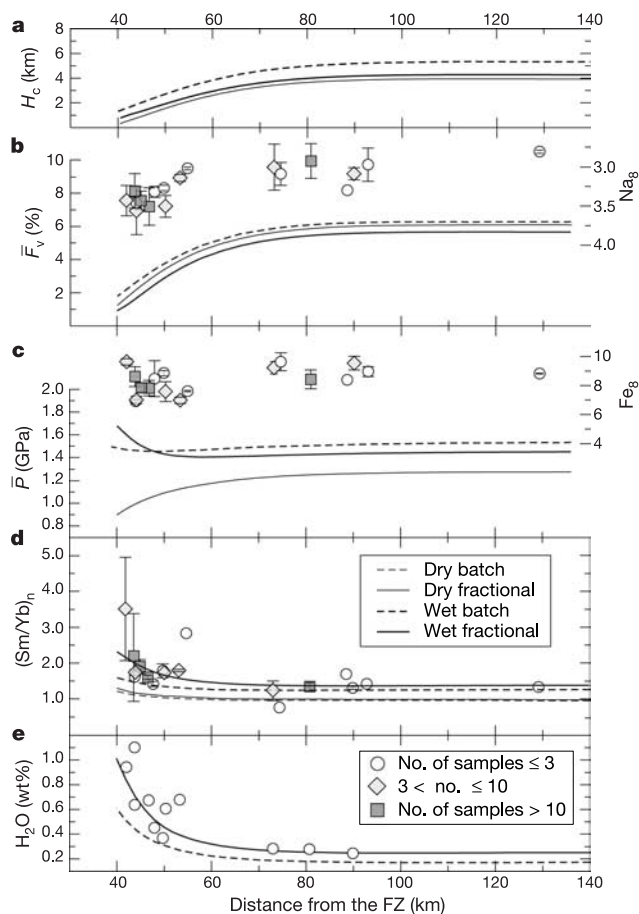


Figure 4 Relationships between melt parameters predicted for MOR melting regimes and values obtained from the basalts sampled along the ERRS axis (Supplementary Table). **a**, Crustal thickness, H_c . **b**, Average degree of melting ($\bar{F}_v$, lines, left axis) and Na_8 (data points, right axis). **c**, Mean pressure of melting ($\bar{P}$, lines, left axis) and Fe_8 (data points, right axis). **d**, Chondrite-normalized³⁰ Sm/Yb ratio, $(\text{Sm/Yb})_n$. The increasing influence of garnet as the ridge–transform intersection is approached is reflected by the increase of $(\text{Sm/Yb})_n$ relative to the source. **e**, Models of water content in the aggregated melt and observed H_2O concentrations in basaltic glasses. Error bars, $\pm 1 \text{ s.d.}$

Methods

Major elements

These were determined using a JEOL JXA 8600 microprobe at IGG-CNR, in Florence. The acceleration voltage was 15 kV, the sample current was 10 nA. The counting times were 40 s for Na and Cl and 10 s for all other elements; the spot size was 10 μm .

H₂O

H₂O content was determined by secondary ion mass spectrometry (SIMS) with a Cameca IMS 4f ion microprobe (at IGG-CNR in Pavia), following a procedure that involves 'energy filtered' secondary ions²⁷ (emission energies in the range 75–125 eV). Under these experimental conditions, the H background, measured on a sample of quartz, is typically 0.009 wt% H₂O. The values for H₂O in the Supplementary Table are the average of three measurements. The accuracy of analysis is estimated to be 10% relative.

REE

REE concentrations were determined with the Pavia ion microprobe. An optimized energy filtering technique was applied to remove complex molecular interferences in the secondary ion mass spectrum. Light-REE-rich basalts were analysed applying a deconvolution filter to the secondary-ion REE mass spectrum in order to reduce residual oxide interferences (that is, BaO on Eu, CeO and NdO on Gd, GdO on Yb, and EuO on Er). Precision of the measurement is of the order of 10% relative, for REE concentrations in the range 0.1–0.7 p.p.m. Below 0.1 p.p.m., precision is mainly limited by (poissonian) counting statistics and falls to ~30% relative. Accuracy is of the same order of precision. The experimental conditions involved a 9.5 nA, ¹⁶O[−] primary ion beam accelerated through −12.5 kV and focused into a spot 10–15 μm in diameter, and energy-filtered (75–125 eV) positive secondary ions detected under an ion image field of 25 μm .

Received 18 June; accepted 30 November 2004; doi:10.1038/nature03264.

- Danyushevsky, L. V. The effect of small amounts of H₂O on crystallisation of mid-ocean ridge and backarc basin magmas. *J. Volcanol. Geotherm. Res.* **110**, 265–280 (2001).
- Michael, P. J. The concentration, behavior and storage of H₂O in the suboceanic upper mantle — Implications for mantle metasomatism. *Geochim. Cosmochim. Acta* **52**, 555–566 (1988).
- Nichols, A. R. L., Carroll, M. R. & Hoskuldsson, A. Is the Iceland hot spot also wet? Evidence from the water contents of undegassed submarine and subglacial pillow basalts. *Earth Planet. Sci. Lett.* **202**, 77–87 (2002).
- Asimow, P. D., Dixon, J. E. & Langmuir, C. H. A hydrous melting and fractionation model for mid-ocean ridge basalts: Application to the Mid-Atlantic Ridge near the Azores. *Geochem. Geophys. Geost.* **5**, Q01E16, doi:10.1029/2003GC000568 (2004).
- Dixon, J. E. & Clague, D. A. Volatiles in basaltic glasses from Loihi Seamount, Hawaii; evidence for a relatively dry plume component. *J. Petrol.* **42**, 627–654 (2001).
- Dixon, J. E. & Stolper, E. M. An experimental study of water and carbon dioxide solubilities in mid-ocean ridge basaltic liquids. 2. Applications to degassing. *J. Petrol.* **36**, 1633–1646 (1995).
- Klein, E. M. & Langmuir, C. H. Global correlations of ocean ridge basalt chemistry with axial depth and crustal thickness. *J. Geophys. Res.* **92**, 8089–8115 (1987).
- Taylor, B. & Martinez, F. Back-arc basin basalt systematics. *Earth Planet. Sci. Lett.* **210**, 481–497 (2003).
- Schilling, J. G. Azores mantle blob: The rare-earth evidence. *Earth Planet. Sci. Lett.* **24**, 103–105 (1975).
- Bonatti, E. Not so hot “hot spots” in the oceanic mantle. *Science* **250**, 107–111 (1990).
- Bonatti, E., Seyler, M. & Sushevskaya, N. A cold suboceanic mantle belt at the Earth's equator. *Science* **261**, 315–320 (1993).
- Schilling, J. G. *et al.* Thermal structure of the mantle beneath the equatorial mid-Atlantic ridge—Inferences from the spatial variation of dredged basalt glass compositions. *J. Geophys. Res.* **100**, 10057–10076 (1995).
- Bonatti, E. *et al.* Steady-state creation of crust-free lithosphere at cold spots in mid-ocean ridges. *Geology* **29**, 979–982 (2001).
- Fox, P. J. & Gallo, D. The tectonics of ridge transform intersections. *Tectonophysics* **104**, 204–242 (1984).
- Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: Implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
- Braun, M. G., Hirth, G. & Parmentier, E. M. The effect of deep damp melting on mantle flow and melt generation beneath mid-ocean ridges. *Earth Planet. Sci. Lett.* **176**, 339–356 (2000).
- Asimow, P. D. & Langmuir, C. H. The importance of water to oceanic mantle melting regimes. *Nature* **421**, 815–820 (2003).
- Katz, R. F., Spiegelman, M. & Langmuir, C. H. A new parameterization of hydrous mantle melting. *Geochem. Geophys. Geost.* **4**, 1073, doi:10.1029/2002GC000433 (2003).
- Gast, P. Trace element fractionations and the origin of tholeiitic and alkaline magma types. *Geochim. Cosmochim. Acta* **32**, 1057–1086 (1968).
- Shen, Y. & Forsyth, D. W. Geochemical constraints on initial and final depths of melting beneath mid-ocean ridges. *J. Geophys. Res.* **100**, 2211–2237 (1995).
- Hellebrand, E., Snow, J. E., Dick, H. J. B. & Hofmann, A. W. Coupled major and trace elements as indicators of the extent of melting in mid-ocean-ridge peridotites. *Nature* **410**, 677–681 (2001).
- Ellam, R. M. Lithospheric thickness as a control on basalt geochemistry. *Geology* **20**, 153–156 (1992).
- McKenzie, D. & O'Nions, R. K. Partial melt distributions from inversion of rare earth element concentrations. *J. Petrol.* **32**, 1021–1091 (1991).
- Hellebrand, E., Snow, J. E., Hoppe, P. & Hofmann, A. W. Garnet-field melting and late-stage refertilization in “residual” abyssal peridotites from the central Indian ridge. *J. Petrol.* **43**, 2305–2338 (2002).
- Christie, D. M., West, B. P., Pyle, D. G. & Hanan, B. B. Chaotic topography, mantle flow and mantle migration in the Australian–Antarctic discordance. *Nature* **394**, 637–644 (1998).
- Michael, P. J. *et al.* Magmatic and magmatic seafloor generation at the ultraslow-spreading Gakkel ridge, Arctic Ocean. *Nature* **423**, 956–961 (2003).

- Ottolini, L., Bottazzi, P., Zanetti, A. & Vannucci, R. Determination of hydrogen in silicates by secondary ion mass spectrometry. *Analyst* **120**, 1309–1314 (1995).
- Phipps Morgan, J. & Forsyth, D. W. Three-dimensional flow and temperature perturbations due to a transform offset: Effects on oceanic crustal and upper mantle structure. *J. Geophys. Res.* **93**, 2955–2966 (1988).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions. *Geophys. Res. Lett.* **21**, 2191–2194 (1994).
- Anders, E. & Grevesse, N. Abundances of the elements: Meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Brunelli for comments, and D. W. Forsyth for providing programs for temperature calculations. This work was supported by the Italian Consiglio Nazionale Ricerche, the US National Science Foundation and the EUROMARGINS programme.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.B. (enrico.bonatti@bo.ismar.cnr.it).

Kin selection and cooperative courtship in wild turkeys

Alan H. Krakauer

Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, California 94720-3160, USA

In the few species of birds in which males form display partnerships to attract females, one male secures most or all of the copulations^{1,2}. This leads to the question of why subordinate males help in the absence of observable reproductive benefits. Hamilton's concept of kin selection³, whereby individuals can benefit indirectly by helping a relative, was a crucial breakthrough for understanding apparently altruistic systems. However in the only direct test of kin selection in coordinated display partnerships, partners were unrelated¹, discounting kin selection as an explanation for the evolution of cooperation. Here I show, using genetic measures of relatedness and reproductive success, that kin selection can explain the evolution of cooperative courtship in wild turkeys. Subordinate (helper) males do not themselves reproduce, but their indirect fitness as calculated by Hamilton's rule^{3,4} more than offsets the cost of helping. This result confirms a textbook example of kin selection² that until now has been controversial⁵ and also extends recent findings^{6–8} of male relatedness on avian leks by quantifying the kin-selected benefits gained by non-reproducing males.

The observation that cooperation in birds typically occurs between relatives⁹ is superficial support for the idea that kin selection is a general explanation for avian helping behaviour. However, most examples of helping or cooperative breeding involve offspring retained in intergenerational family groups^{9,10} in which it is difficult to separate the indirect fitness benefits due to kin selection from benefits due to direct fitness¹¹, even when the dynamics of helping behaviour qualitatively fits the predictions of Hamilton's rule¹². Species with aggregated male displays are therefore valuable for studying kin selection because it is possible to isolate the role of indirect fitness in the absence of direct benefits stemming from delayed dispersal. Surprisingly, the role of kin selection on leks has only recently been proposed¹³, and kin associations of displaying males have now been demonstrated for several species including grouse⁶, peafowl⁷ and manakins⁸. This

kinship facilitates indirect fitness benefits for low-ranked males because female visitation often increases at larger leks, thereby explaining why a male might settle at one lek rather than another^{13,14}. No empirical study has yet quantified this kin-selected benefit relative to alternative benefits such as low frequencies of matings or future inheritance of dominant positions.

Wild turkeys (*Meleagris gallopavo*) are among the few species of birds to form male display partnerships within larger display aggregations. Male turkeys form coalitions of two to four same-aged males that court females and defend those females against other groups and solitary males. A previous observational study² found that only one male in each coalition mates. The authors of that study believed that coalitions were composed of brothers (sibling nestmates), thereby providing the relatedness necessary to accrue indirect fitness benefits. Although this work was published long before molecular techniques were available to measure relatedness and reproductive success directly and has been questioned on other grounds⁵, the story remains a well-cited example of kin selection^{15,16}.

Here I show that kin-selected indirect fitness benefits do in fact explain cooperative courtship in wild turkeys. Support for kin selection requires the following three conditions to be met: first, that dominant and subordinate males are related; second, that there is a measurable benefit to the dominant male due to the help of the subordinate; and last, that Hamilton's rule ($rB - C < 0$) must be met, the indirect benefit to the subordinate (rB) outweighing the cost of helping instead of attempting to breed independently (C) (refs 3, 4; Table 1).

First, coalitions are clearly composed of relatives, as shown by the similarity of microsatellite genotypes between dominant and subordinate males (Fig. 1). Six coalitions (five pairs and one four-member group for a total of eight dominant-subordinate male dyads) combined for a mean ($\pm$ s.e.m.) coalition relatedness, r , of 0.42 ± 0.07 , significantly higher than r of males drawn randomly from the population (randomization test, 1,000 iterations, $P < 0.001$). The relatedness within coalitions is equivalent to the mean r calculated for two groups of known genealogical relationship with expected $r = 0.5$ based on pedigree (full siblings, $r = 0.52 \pm 0.05$, $n = 10$; mothers and their offspring, $r = 0.46 \pm 0.03$, $n = 12$, analysis of variance $F_{(2,27)} = 0.947$, $P = 0.40$). It was not possible to test specific genealogical hypotheses for individual coalitions because of an insufficient number of loci¹⁷.

Second, the help provided by subordinate males increases the reproductive success of dominant males compared with non-cooperating solitary males (Fig. 2). Dominant males mated with more females ($\chi^2_{(1)} = 9.0$, $P < 0.005$) and fathered more offspring than solitary males ($\chi^2_{(1)} = 58.3$, $P < 0.001$). The benefit B , as calculated by the difference between the mean fitness of dominant males and the mean fitness of solitary males, was 6.1 offspring per

male (Table 1). This value is an underestimate to the extent that high relatedness makes it more difficult to assign paternity to males in coalitions (see Methods).

An alternative hypothesis is that the benefit is due to differences in individual quality of dominant males rather than the help provided by their subordinate partner(s). If this were true, one would predict a difference in the distributions of reproductive success between dominant and solitary males, and that there would be little or no difference between the success of presumed high-quality males that successfully reproduce, whether they are in a dominant member of a coalition or display as a solitary individual. Contrary to these predictions, both dominants and solitary males show bimodal distributions of reproductive success, with $n = 3$ dominant males and $n = 10$ solitary males not reproducing at all. When only the presumably high-quality males that reproduce are used to calculate the average fitness of dominant and solitary males, the dominant males father significantly more offspring (49) than do solitary males (13) (Mann-Whitney $U = 0.5$, $P = 0.026$, $n = 4$ for both groups), and the benefit B increases to 9.0 offspring per male rather than decreases (Table 1). Thus kin selection seems to best explain the pattern of relatedness and distribution of reproductive success; the important issue of individual male quality that has been studied in captive settings^{18,19} remains to be integrated into the complex mating system of free-living turkeys.

Finally, Hamilton's rule can be evaluated by assuming the cost of helping, C , for subordinate male turkeys is equal to the average fitness of non-cooperative solitary males (0.9 offspring per male). With this assumption, the net benefit to helping is +1.7 offspring per male (Table 1), indicating a clear selective benefit to cooperation for subordinate males. Similar results are obtained if presumably low-quality (non-reproducing) males are excluded from calculations (Table 1). This benefit is calculated on the assumption that all coalitions are dyads. One coalition was initially observed to contain three males and another four males, although these both were reduced to pairs during their first season, presumably by hunting or natural predation events. The fitness benefits for third- or lower-ranked males would require further assumptions, including size-specific coalition productivity, which I could not calculate on the basis of my limited sample size.

The minimum level of relatedness necessary to offset a subordinate's loss of independent reproductive opportunities can be

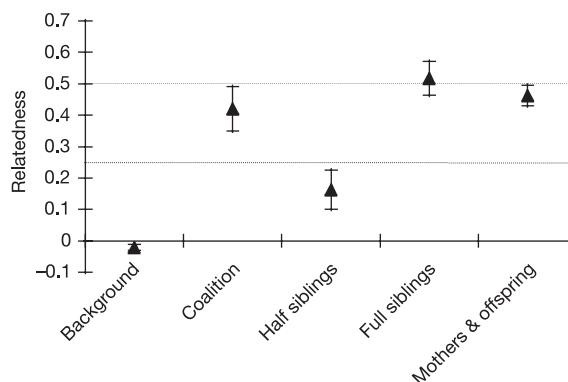


Figure 1 Relatedness values calculated from microsatellite genotypes. The graph shows the background relatedness for all adult males and the relatedness of subordinate males to their dominant partner ('Coalition'). Dotted lines represent expected values for full and half siblings; unrelated individuals should have $r = 0$. Relatedness of three known genealogical relationships, based on parentage analysis, indicate that calculated r values are concordant with their predicted values based on pedigree. These groups were: half siblings sharing either a mother or a father, full sibling nest mates and mothers and their offspring. Values are means $\pm$ s.e.m. Sample sizes, from left to right, are $n = 1,250$, 8, 10, 10 and 12.

Table 1 Calculation of Hamilton's rule, $rB - C < 0$

Variable	Description	Calculation	Value*
r	Coefficient of relatedness	Mean pairwise relatedness of subordinates to their dominant display partner	0.42
$B^\dagger$	Benefit to dominant	(No. of offspring per dominant male) – (no. of offspring per solo male)	6.1 (9.0)
$C^\dagger$	Cost to subordinate	(No. of offspring per solo male) – (no. of offspring per subordinate male)	0.9 (2.3)
	Net benefit ‡	$rB - C$	+1.7 (1.5)

Dominant, solo and subordinate refer to dominant coalition males, solitary non-cooperating males and subordinate male helpers, respectively.

*Values in parentheses exclude non-reproducing males from mean fitness calculations.

‡ In units of offspring per male.

calculated by setting the indirect benefit, rB , equal to the cost C and solving for r . On the basis of the values in Table 1, helping behaviour should be observed as long as coalitions are related at a level of $r > 0.15$. Given that half-brothers have an expected r value of 0.25, processes that produce half-siblings such as multiple paternity, quasiparasitism or crècheing of broods fathered by the same male should not reduce relatedness among brood-mates to the point in which cooperative behaviour is no longer favoured. This conclusion is robust to an almost 50% underestimate in relative fitness of solitary males.

How do turkeys compare to other bird species that form male display partnerships? In contrast to reproductive 'sharing' by satellite males in ruff (*Philomachus pugnax*) partnerships²⁰, I found no evidence of direct reproduction by subordinate male turkeys (Fig. 2). None of eight marked subordinate males fathered any offspring during my study, compared with four out of seven dominant members of coalitions (Fisher exact test, $P = 0.026$).

In cooperative partnerships of *Chiroxiphia* manakins, subordinate males seem to benefit through an increased likelihood of future inheritance of a display perch¹. Unlike *Chiroxiphia*, wild turkey coalitions do not act as social queues because coalitions change only through attrition. Coalitions form before adulthood (three coalitions were marked as 1-year-old subadults); furthermore, no solitary displaying male was observed to later join a coalition ($n = 14$ males and 24 male-years). This trend cannot be explained by solitary males joining distant coalitions outside the study area, because coalitions and solitary males were observed at similar rates (see Supplementary Table 1). All changes in coalition membership observed were losses rather than gains of individuals ($n = 7$ coalitions across 11 coalition-years; 6 losses, 0 gains, Sign test

$P < 0.04$). This pattern indicates that if the dominant male disappears from a cooperative pair, the subordinate is left as a solitary male and does not attract a new display partner. Finally, turkeys do not defend territories either during the breeding season, when several male groups may court a given flock of females, or outside the breeding season, when males are highly social^{2,21}. Thus future resource or territory inheritance cannot account for subordinate male cooperation.

Because subordinate males acquire large indirect fitness benefits, do not themselves gain direct reproduction and are unlikely to increase their future mating opportunities, kin selection seems to provide the best explanation for the evolution of cooperative behaviour in wild turkeys. By contrasting these results with the patterns described for *Chiroxiphia* manakins¹ and ruffs²⁰, it is evident that although these species independently evolved cooperative courtship as a solution to intensely competitive mating systems, the exact form of fitness benefit maintaining subordinate cooperation can differ greatly. □

Methods

Field methods

From 1999 to 2004 I studied an introduced population of *M. gallopavo* at the Hastings Natural History Reservation in Carmel Valley, Monterey County, California. I captured 126 immature and adult turkeys (51 males, 75 females) by using walk-in traps or drop nets. About 100 μ l of blood was taken from the wing vein and stored in blood storage buffer. Adults were tagged with uniquely numbered patagial wing tags, and a subset ($n = 8$ males, $n = 68$ females) were outfitted with backpack-style radiotransmitters. About 50% of the population was marked in any given year. With one or two field assistants each year, I attempted to relocate radio-tagged birds visually at least twice a week in January, February and June, and daily from March to May, to identify female nesting attempts and to observe courtship behaviour and associations. In addition, we regularly drove along about 20 km of roads in and around the reserve, and during the breeding season we hiked at least a 2-km loop at a nearby ranch to search for turkeys without radios. Genetic samples from offspring ($n = 325$) were collected by a combination of capturing flightless young soon after they had hatched, salvaging from failed or abandoned nests, and collecting early nests to incubate and sample eggs. These samples were stored in one or more of the following: blood storage buffer, dimethyl sulphoxide or 100% ethanol. All procedures were approved by the University of California, Berkeley, and the California Department of Fish and Game.

Behavioural definitions

Coalitions were defined as adult males in their third year (2 years old) or older that were seen displaying to females at least twice while within 2 m of each other. Solitary males were males that never met this criterion and either were observed at least twice displaying alone or showed patterns of association that precluded them from having a specific partnership with another male. Within a coalition, the dominant male was the one that performed most of the full strut (stereotyped pulmonary puff) displays. Data on individual coalitions and solitary males are provided in Supplementary Table 1.

DNA extraction and genotyping

DNA was extracted from samples using DNEasy tissue extraction kits (Qiagen) then diluted to a concentration of 20 μ g ml⁻¹. All individuals were then genotyped at ten microsatellite loci identified from previous studies of wild or domestic turkeys²²⁻²⁴. Using GENEPOP²⁵, I determined that these loci were unlinked. Additional details such as reaction conditions and allelic diversity are provided in Supplementary Table 2. One primer of each primer pair was fluorescently labelled and loci were multiplexed on an ABI 3730 automated sequencer. Polymerase chain reaction products were run on 96-well plates that contained one negative control and two positive controls (the same two individuals were included with every run). All adults were genotyped at least twice, and genotypes were more than 0.999 complete. Offspring genotypes were 0.98 complete; individuals were rerun if allelic calls were questionable or if they did not match the genotype of the incubating female. Some level of mismatching is to be expected because of both marker mutation and nest parasitism by other females.

Analysis of genetic data

Relatedness was calculated with RELATEDNESS 5.0 (ref. 26). The background allele frequencies were defined as the allele frequencies for adults only. Relatedness values were then calculated for all pairs of males, and coalition values were compared with 1,000 sets of eight randomly selected values. Sets of full siblings, half siblings and mother-offspring pairs were identified during subsequent paternity analyses (see below); r values were calculated for these known genealogical relationships to confirm that the relatedness estimates generated from microsatellite genotypes corresponded to those predicted by pedigree.

Reproductive success was determined by assigning parentage to sampled offspring. Maternity of an offspring was assigned to the female incubating that nest if she had no more than one locus mismatching the offspring. Paternity was calculated by a combination of maximum likelihood assignment with CERVUS 2.0 (ref. 27) and

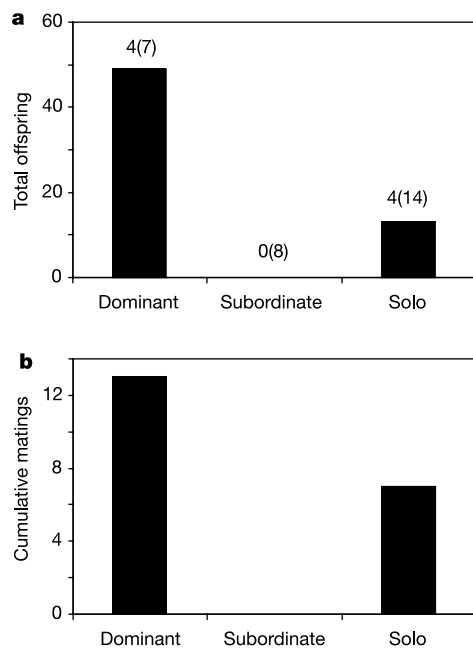


Figure 2 Reproductive success of the three male display strategies in wild turkeys. Males are classified as either dominant coalition member, subordinate coalition helper or non-cooperating solitary male. Reproductive success is shown as total number of offspring (a) and total number of mates (b). The number of males assigned paternity, with total sampled males in parentheses, are indicated above the bars in a. Subordinate males fathered no offspring. Dominant males fathered significantly more offspring than solitary males ($\chi^2_{(1)} = 58.3$, $P < 0.001$) and mated with significantly more females ($\chi^2_{(1)} = 9.0$, $P < 0.005$).

genotypic exclusion, meaning that a male could be assigned paternity only if he met the strict 95% assignment level and was the only perfect genotypic match among sampled males. Details of the CERVUS analysis are given in Supplementary Table 3.

This set of conservative criteria lead to the assignment of 75 of 325 offspring to a known, sampled male. Although many of the unassigned offspring were probably fathered by unsampled males, others were fathered by known males but could not be assigned to them. CERVUS generates a test statistic (Δ) based on the difference in LOD scores (cumulative log-likelihood ratio of parentage compared with non-parentage) between the two most likely sampled males. The program then compares this value with a user-defined critical value based on the percentage of simulations (here, 95%) that correctly assigned an offspring to the actual parent. The more genotypically similar that two candidate parents are, the more likely they are to have similar LOD scores and therefore to generate a smaller Δ score. Given the wild turkey's unique kin structure, coalition males were handicapped by necessarily having close relatives among the set of candidate males.

Received 20 November; accepted 30 December 2004; doi:10.1038/nature03325.

- McDonald, D. B. & Potts, W. K. Cooperative display and relatedness among males in a lek-mating bird. *Science* **266**, 1030–1032 (1994).
- Watts, C. R. & Stokes, A. W. The social order of turkeys. *Sci. Am.* **224**, 112–118 (1971).
- Hamilton, W. D. The genetical theory of social behavior. I and II. *J. Theor. Biol.* **7**, 1–52 (1964).
- Lucas, J. R., Creel, S. R. & Waser, P. M. How to measure inclusive fitness, revisited. *Anim. Behav.* **51**, 225–228 (1996).
- Balphy, D. F., Innes, G. S. & Balphy, M. H. Kin selection in turkeys: a critical assessment. *Auk* **97**, 854–860 (1980).
- Höglund, J., Alatalo, R. V., Lundberg, A., Rintamäki, P. K. & Lindell, J. Microsatellite markers reveal the potential for kin selection on black grouse leks. *Proc. R. Soc. Lond. B* **266**, 813–816 (1999).
- Petrie, M., Krupa, A. & Burke, T. Peacocks lek with relatives even in the absence of social and environmental cues. *Nature* **401**, 155–157 (1999).
- Shorey, L., Piattney, S., Stone, J. & Höglund, J. Fine-scale genetic structure on *Manacus manacus* leks. *Nature* **408**, 352–353 (2000).
- Brown, J. L. *Helping and Communal Breeding in Birds* (Princeton Univ. Press, Princeton, 1987).
- Emlen, S. T. Benefits, constraints and the evolution of the family. *Trends Ecol. Evol.* **9**, 282–285 (1994).
- Clutton-Brock, T. H. Breeding together—kin selection and mutualism in cooperative vertebrates. *Science* **296**, 69–72 (2002).
- Griffin, A. S. & West, S. A. Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. *Science* **302**, 634–636 (2003).
- Kokko, H. & Linström, J. Kin selection and the evolution of leks: whose success do young males maximize? *Proc. R. Soc. Lond. B* **263**, 919–923 (1996).
- Höglund, J. Lek-kin in birds—provoking theory and surprising new results. *Ann. Zool. Fenn.* **40**, 249–253 (2003).
- Wilson, E. O. *Sociobiology: the New Synthesis* (Harvard Univ. Press, Boston, Massachusetts, 1975).
- Goodenough, J., McGuire, B. & Wallace, R. A. *Perspectives on Animal Behavior* 2nd edn (Wiley, New York, 2001).
- Blouin, M. S. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends Ecol. Evol.* **18**, 503–511 (2003).
- Buchholz, R. Female choice, parasite load, and male ornamentation in wild turkeys. *Anim. Behav.* **50**, 929–943 (1995).
- Buchholz, R. Male dominance and variation in fleshy head ornamentation in wild turkeys. *J. Avian Biol.* **28**, 223–230 (1997).
- Lank, D. B. et al. High frequency of polyandry in a lek mating system. *Behav. Ecol.* **13**, 209–215 (2002).
- Healy, W. M. *The Wild Turkey: Biology and Management* (ed. Dickson, J. D.) 46–65 (Stackpole, Mechanicsburg, Pennsylvania, 1992).
- Donoghue, A. M., Sonstegard, T. S., King, L. M., Smith, E. J. & Burt, D. W. Turkey sperm mobility influences paternity in the context of competitive fertilization. *Biol. Reprod.* **61**, 422–427 (1999).
- Reed, K. M., Roberts, M. C., Murtaugh, J., Beattie, C. W. & Alexander, L. J. Eight new dinucleotide microsatellite loci in turkey (*Meleagris gallopavo*). *Anim. Genet.* **31**, 140 (2000).
- Mock, K. E., Theimer, T. C., Rhodes, O. E., Greenberg, D. L. & Keim, P. Genetic variation across the historical range of the wild turkey (*Meleagris gallopavo*). *Mol. Ecol.* **11**, 643–657 (2002).
- Raymond, M. & Rousset, F. GENEPOP (version 3.1b). An updated version of GENEPOP version 1.2: population genetics software for exact tests and ecumenicism. *J. Hered.* **86**, 248–249 (1997).
- Queller, D. C. & Goodnight, K. F. Estimating relatedness using genetic markers. *Evolution* **43**, 258–275 (1989).
- Marshall, T. C., Slate, J., Kruuk, L. E. B. & Pemberton, J. M. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7**, 639–655 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements I thank E. DuVal, E. Lacey and especially W. Koenig and M. Hauber for comments on the manuscript; J. Dickinson, S. Beissinger, B. Jones, B. Slikas and N. Johnson for training and discussion; B. Davis, A. Hsieh, F. Aguilar, C. Miller, J. Leyhe, R. Orben, L. Robinson, M. Nguyen, B. Loui, M. Nasiri and C. Chu and others for field and laboratory assistance; and P. Kephart for research access to Rana Creek Ranch. My research was funded by an NSF graduate fellowship and dissertation improvement grant, as well as the Animal Behavior Society, American Ornithologists' Union, Sigma Xi, and support from both Integrative Biology and the Museum of Vertebrate Zoology at the University of California, Berkeley.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to the author (krakauer@berkeley.edu).

Disruptive coloration and background pattern matching

Innes C. Cuthill¹, Martin Stevens¹, Jenna Sheppard¹, Tracey Maddocks¹, C. Alejandro Párraga² & Tom S. Troscianko²

¹School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG, UK

²Department of Experimental Psychology, University of Bristol, Woodland Road, Bristol BS8 1TN, UK

Effective camouflage renders a target indistinguishable from irrelevant background objects. Two interrelated but logically distinct mechanisms for this are background pattern matching (crypsis^{1,2}) and disruptive coloration: in the former, the animal's colours are a random sample of the background^{1,2}; in the latter, bold contrasting colours on the animal's periphery break up its outline. The latter has long been proposed as an explanation for some apparently conspicuous coloration in animals^{3,4}, and is standard textbook material. Surprisingly, only one quantitative test⁵ of the theory exists, and one experimental test of its effectiveness against non-human predators⁶. Here we test two key predictions: that patterns on the body's outline should be particularly effective in promoting concealment and that highly contrasting colours should enhance this disruptive effect. Artificial moth-like targets were exposed to bird predation in the field, with the experimental colour patterns on the 'wings' and a dead mealworm as the edible 'body'. Survival analysis supported the predictions, indicating that disruptive coloration is an effective means of camouflage, above and beyond background pattern matching.

The pioneers of modern military camouflage were both artists and keen observers of nature⁷. For example, the work of Thayer, who proposed the theory of countershading⁸ and developed Bates's ideas on disruptive coloration in animals³, was influential in persuading

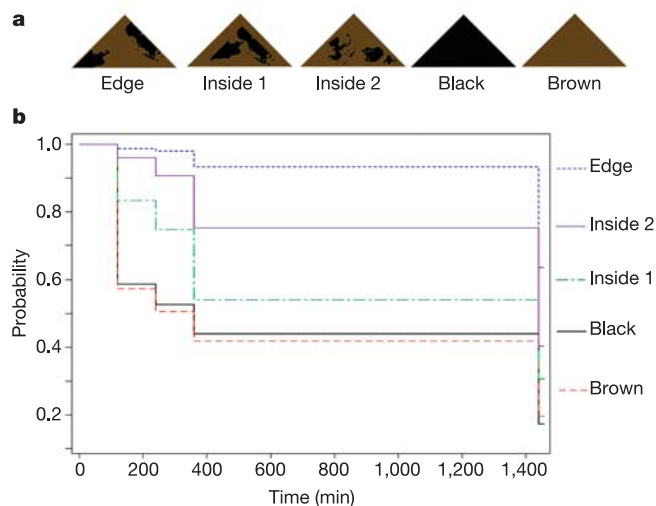


Figure 1 Patterns placed on the body's outline enhance survival. **a**, Examples of 'moth' targets in experiment 1; **b**, survival curves. The differences between treatments were significant (Wald = 138.92, d.f. = 4, $P < 0.001$) in the order Edge > Inside 2 (Wald = 16.03, d.f. = 1, $P < 0.001$) > Inside 1 (Wald = 11.01, d.f. = 1, $P = 0.001$) > Black or Brown (Inside 1 versus Black, Wald = 13.33, d.f. = 1, $P < 0.001$; Inside 1 versus Brown, Wald = 13.11, d.f. = 1, $P < 0.001$); there was no difference between the latter monochrome treatments (Wald = 0.00, d.f. = 1, $P = 0.992$).

the US government to form a special camouflage unit during the First World War (ref. 7). Thayer's theory of 'ruptive' coloration, that contrasting colours on an object help to break up its outline, is such a central feature of military camouflage—and similar patterns in the animal kingdom seem so obviously designed to fulfil the same role⁴—that it is easy to assume that what deceives humans deceives other animals. Yet it is unwise to extrapolate from human perception to that of other animals^{9–11}, and there are many other explanations for colour patterns presumed by Thayer and successors to be camouflage^{3,4,12,13}. Before accepting that disruptive coloration has a role in anti-predator defence, two conditions must hold: that the distribution of colour patterns matches that predicted by the theory, and that these distributions reduce the detectability of prey to predators. There has only been one test of each condition. In one⁵, the spots on a marine isopod were shown to touch the body outline more often than predicted by background matching, which is consistent with the theory of disruptive coloration. In the other⁶, there was no effect on survival of experimentally removing the wing-stripes of a nymphalid butterfly that is highly palatable to birds, a finding inconsistent with the theory, although the methods might have unintentionally made the butterflies more similar to a co-occurring unpalatable species¹⁴ or altered their palatability directly. Clearly there is a pressing need for further empirical research before we can accept what has been described⁴ as "certainly the most important set of principles relating to concealment".

We tested two predictions⁵ arising from previous work^{3,4}: first, that patterns on the body's edge should be more effective than equivalent patterns placed randomly; second, that highly contrasting colours should be more disruptive than those of low contrast. In each case, provided that the colours and patterns on the prey were equally common in the background, the theory of background pattern matching^{1,2,15} would predict no difference in the effective-

ness of the camouflage. Our artificial targets (see Methods), with their coloured 'wings' and edible 'bodies', were not designed to mimic any real lepidopteran, and the oak tree trunks on which they were pinned were merely a convenient complexly patterned substrate against which birds might detect prey. Thus, our experiments are best thought of as 'field psychophysics' rather than a study of moth predation risk.

In experiment 1, targets were dark brown with black markings designed, with regard to bird vision, to match the real patterns of dark and light on heavily ridged, mature, oak bark. There were five treatments (Fig. 1): markings overlapping the edges of the 'wings' ('Edge'), the exact same markings displaced inwards so that no edges were overlapped ('Inside 1'), other randomly selected markings placed so as not to overlap edges ('Inside 2'), monochrome brown, and monochrome black. The three bicoloured treatments all possessed life-sized pattern elements randomly sampled from the background, and so should have been equally cryptic in terms of background pattern matching (and better camouflaged than monochrome brown or black). Similarly, no difference would be predicted if bicoloured targets gained a crypsis advantage because, when viewed from a distance, predators would not be able to discriminate between the two colours and so would see a spatially averaged dark brown. Only the theory of disruptive coloration predicted that treatment Edge should survive better than the other bicoloured treatments. This prediction was fulfilled (Fig. 1). Treatment Inside 2 was included because of the possibility that moving the pattern elements present in treatment Edge from the periphery of the 'wings', to form treatment Inside 1, created pattern elements with straight lines that themselves could have enhanced conspicuousness. This indeed seemed to be so, because treatment Inside 2 survived better than Inside 1, which lacked these straight edges to the pattern elements (Fig. 1). The inwards displacement of pattern elements in Inside 1 also tended to enhance the outline of these targets, thus having the opposite effect to disruptive coloration. Nevertheless, all bicoloured treatments survived better than monochrome black or brown, indicating that background pattern matching was, as expected, itself effective as camouflage (Fig. 1).

Experiment 2 had six treatments: the 2 × 2 combination of bicoloured patterns with high or low contrast, placed as in experiment 1's treatment Edge or Inside 2, plus two monochrome treatments that were the average colour of either the high-contrast or the low-contrast colour pairs. As uniquely predicted by the theory of disruptive coloration, the high-contrast-edge treatment survived best (Fig. 2), with high contrast providing minimal benefit in non-disruptive 'Inside' treatments. The results apply to the conditions pertaining in our study (for example winter, and a given habitat type); the extent to which disruptive patterns provide a general advantage over simple crypsis, with different background types (for example varying spatial and/or chromatic complexity) or different light environments (for example direct or diffuse lighting) therefore awaits further experimentation. Nevertheless, our results provide the strongest support so far for the effectiveness of disruptive patterns against birds, the most commonly invoked visual predators shaping the evolution of protective coloration in insects. □

Methods

'Prey' were dead (frozen overnight at -80 °C, then thawed) mealworms (*Tenebrio molitor* larvae) pinned onto coloured paper triangles 50 mm wide by 25 mm high. These were pinned onto oak trees in the mixed deciduous Leigh Woods National Nature Reserve, North Somerset, UK (2° 38.6' W, 51° 27.8' N) and their 'survival' was checked at about 2, 4, 6 and 24 h. Birds took all or most of the mealworm, spiders sucked fluids out, leaving a hollow exoskeleton, and slugs left slime trails; predation in the latter two categories, complete disappearance of a target, or survival to 24 h, were treated as 'censored' values in survival analysis. Both experiments had randomized block designs with ten replicate blocks, run in different areas of the wood on different dates between October 2003 and March 2004. Each block had 75 (experiment 1; 15 per treatment) or 84 (experiment 2; 14 per treatment) targets in a nonlinear transect of about 1.5 km × 20 m (targets on less than 5% of the available trees in each transect). Treatments were randomly allocated to trees,

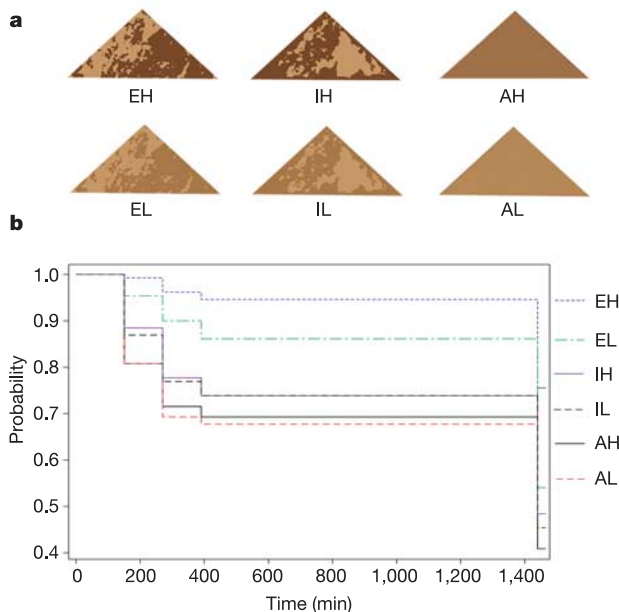


Figure 2 High-contrast disruptive patterns enhance survival. **a**, Examples of 'moth' targets in experiment 2; **b**, survival curves. The differences between treatments were significant (Wald = 62.26, d.f. = 5, $P < 0.001$) in the order Edge-high-contrast (EH) > Edge-low-contrast (EL; Wald = 15.31, d.f. = 1, $P < 0.001$) > Inside-high-contrast (IH) (Wald = 5.20, d.f. = 1, $P = 0.023$) = Inside-low-contrast (IL) (Wald = 0.00, d.f. = 1, $P = 0.952$) = Average-low-contrast (AL) (Wald = 1.68, d.f. = 1, $P < 0.195$) = Average-high-contrast (AH) (Wald = 0.00, d.f. = 1, $P = 0.951$).

subject to the constraints that no lichen covered the trunk and no young trees with a trunk circumference less than 0.9 m were used. Colour matches of treatments to natural bark were verified by spectrophotometry of stimuli and bark, followed by modelling of predicted photon catches¹⁶ of a typical passerine bird, the blue tit's (*Parus caeruleus*) single cone photoreceptors¹⁷, with irradiance spectra from overcast skies in the study site. Our acceptance criterion was simply that cone captures for the experimental stimuli fell within the measured range of those for oak bark.

Experiment 1 used black patterns printed onto dark brown card. Patterns were samples of digital photos of the oak trees at 1:1 reproduction, converted using ImageJ¹⁸ to greyscale and thresholded at 50% to binary (black/white) images to provide, when printed onto brown card, bark-like brown/black spatial variation (Fig. 1). Different samples, from different trees, were used for each replicate target.

Experiment 2 used bicoloured targets printed onto waterproof paper (Hewlett Packard Laserjet Tough Paper) with a Hewlett Packard Colour Laserjet 2500 (600 dots per inch) printer, with colour pairs chosen to have either high or low contrast. Colours were chosen from frequency distributions of the eight-bit RGB (red, green, blue) values from digital photographs of the oak trees in the study site, reduced to 16 bins in each colour channel. Photos (about 267 mm × 200 mm; 2,560 pixels × 1,920 pixels) were taken with a Nikon Coolpix 5700 camera, calibrated¹⁹ to linearize the relationship between radiance and the greyscale in each colour channel, and saved as uncompressed TIFF files. Digital photographs lack ultraviolet information that birds can see²⁰, but lichen-free oak bark reflects negligible ultraviolet²¹. Even a properly calibrated RGB image does not precisely simulate the avian-perceived colour of many natural objects, owing to differences in the spectral sensitivity of bird long-wave, medium-wave and short-wave cones compared with human cones²². However, because our treatments varied only in relative colour contrast, any error associated with this method was considered minor, an assumption verified retrospectively by spectrophotometry and colour-space modelling. We chose colour pairs from the eight most frequent RGB triplets in the bark photos as follows: a 'background' colour, then a triplet that was similar to the background (low contrast), and one that differed markedly (high contrast). The major difference between colours was in overall brightness not hue, but we could not systematically vary only one colour dimension within the available common bark colours. Sample numbers of background and contrasting colours were balanced for which was darker/lighter, and so there were no significant differences between bicoloured treatments in the brightest or darkest colour or average colour (analyses of variance on RGB sums and all possible ratios; $P > 0.9$). Monochrome treatments were also created as the means of the respective R, G and B values of the two colours in bicoloured high-contrast and low-contrast treatments. Different colour pairs and patterns, from different trees, were used for each replicate target.

Survival analysis was by Cox regression^{23,24} with the factors treatment and block. Cox regression assumes that all survival functions have the same shape; this proportional hazards assumption was checked by plotting partial residuals against ranked survival times²⁴. There were significant block effects in both experiments (in experiment 1, Wald = 121.78, d.f. = 9, $P < 0.001$; in experiment 2, Wald = 271.50, d.f. = 9, $P < 0.001$), reflecting differences in average predation rates in different parts of the woods on different dates, but this was not relevant to our hypotheses.

Received 23 September; accepted 30 December 2004; doi:10.1038/nature03312.

1. Endler, J. A. Progressive background in moths, and a quantitative measure of crypsis. *Biol. J. Linn. Soc.* **22**, 187–231 (1984).
2. Endler, J. A. An overview of the relationships between mimicry and crypsis. *Biol. J. Linn. Soc.* **16**, 25–31 (1981).
3. Thayer, G. H. *Concealing Coloration in the Animal Kingdom; An Exposition of the Laws of Disguise through Color and Pattern; Being a Summary of Abbott H. Thayer's Discoveries* (Macmillan, New York, 1909).
4. Cott, H. B. *Adaptive Coloration in Animals* (Methuen, London, 1940).
5. Merilaita, S. Crypsis through disruptive coloration in an isopod. *Proc. R. Soc. Lond. B* **265**, 1059–1064 (1998).
6. Silberglied, R. E., Aiello, A. & Windsor, D. M. Disruptive coloration in butterflies - lack of support in *Anartia fatima*. *Science* **209**, 617–619 (1980).
7. Behrens, R. R. *False Colors: Art, Design and Modern Camouflage* (Bobolink, Dysart, Iowa, 2002).
8. Thayer, A. H. The law which underlies protective coloration. *Auk* **13**, 124–129 (1896).
9. Endler, J. A. On the measurement and classification of colour in studies of animal colour patterns. *Biol. J. Linn. Soc.* **41**, 315–352 (1990).
10. Endler, J. A. A predator's view of animal color patterns. *Evol. Biol.* **11**, 319–364 (1978).
11. Bennett, A. T. D., Cuthill, I. C. & Norris, K. J. Sexual selection and the mismeasure of color. *Am. Nat.* **144**, 848–860 (1994).
12. Kiltie, R. A. Countershading: universally deceptive or deceptively universal? *Trends Ecol. Evol.* **3**, 21–23 (1988).
13. Ruxton, G. D., Speed, M. P. & Kelly, D. J. What, if anything, is the adaptive function of countershading? *Anim. Behav.* **68**, 445–451 (2004).
14. Waldbauer, G. P. & Sternburg, J. G. A pitfall in using painted insects in studies of protective coloration. *Evolution* **37**, 1085–1086 (1983).
15. Merilaita, S., Tuomi, J. & Jormalainen, V. Optimization of cryptic coloration in heterogeneous habitats. *Biol. J. Linn. Soc.* **67**, 151–161 (1999).
16. Maddocks, S. A., Church, S. C. & Cuthill, I. C. The effects of the light environment on prey choice by zebra finches. *J. Exp. Biol.* **204**, 2509–2515 (2001).
17. Hart, N. S., Partridge, J. C., Cuthill, I. C. & Bennett, A. T. D. Visual pigments, oil droplets, ocular media and cone photoreceptor distribution in two species of passerine: the blue tit (*Parus caeruleus* L.) and the blackbird (*Turdus merula* L.). *J. Comp. Physiol. [A]* **186**, 375–387 (2000).
18. Rasband, W. *ImageJ* v. 1.30 (<http://rsb.info.nih.gov/ij/docs/>, National Institutes of Health, USA, 2003).
19. Parraga, C. A., Troscianko, T. & Tolhurst, D. J. Spatiochromatic properties of natural images and human vision. *Curr. Biol.* **12**, 483–487 (2002).
20. Cuthill, I. C. *et al.* Ultraviolet vision in birds. *Adv. Stud. Behav.* **29**, 159–214 (2000).

21. Majerus, M. E. N., Brunton, C. F. A. & Stalker, J. A bird's eye view of the peppered moth. *J. Evol. Biol.* **13**, 155–159 (2000).
22. Cuthill, I. C. *et al.* Avian colour vision and avian video playback experiments. *Acta Ethol.* **3**, 29–37 (2000).
23. Cox, D. R. Regression models and life-tables. *J. R. Stat. Soc. B* **34**, 187–220 (1972).
24. *SPSS for Windows Release 9.0* (SPSS Inc., Chicago, 2003).

Acknowledgements We thank J. Endler for suggestions. The research was supported by a BBSRC grant to I.C.C., T.S.T. and J. C. Partridge.

Authors' contributions I.C.C. designed the experiments and stimuli; M.S., J.S., T.M. and I.C.C. performed the experiments; A.P. wrote the programs for colour analysis and camera calibration; T.S.T. advised on design and colour modelling.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.C. (i.cuthill@bristol.ac.uk).

An anaerobic mitochondrion that produces hydrogen

Brigitte Boxma^{1*}, Rob M. de Graaf^{1*}, Georg W. M. van der Staay^{1*}, Theo A. van Alen¹, Guenola Ricard², Toni Gabaldón², Angela H. A. M. van Hoek^{1†}, Seung Yeo Moon-van der Staay¹, Werner J. H. Koopman³, Jaap J. van Hellemond⁴, Aloysius G. M. Tielens⁴, Thorsten Friedrich⁵, Marten Veenhuis⁶, Martijn A. Huynen² & Johannes H. P. Hackstein¹

¹Department of Evolutionary Microbiology, Faculty of Science, Radboud University Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands

²Centre for Molecular and Biomolecular Informatics,

³Microscopical Imaging Centre and Department of Biochemistry, Nijmegen Centre of Molecular Life Sciences (NCMLS), Radboud University Nijmegen Medical Centre, NL-6500 HB Nijmegen, The Netherlands

⁴Department of Biochemistry and Cell Biology, Faculty of Veterinary Medicine, Utrecht University, PO Box 80176, NL-3508 TD Utrecht, The Netherlands

⁵Albert-Ludwigs-Universität, Institut für Organische Chemie und Biochemie, Albertstrasse 21, D-79104 Freiburg i. Br., Germany

⁶Department of Eukaryotic Microbiology, Groningen University, PO Box 14, NL-9750 AA Haren, The Netherlands

* These authors contributed equally to this work

† Present address: RIKILT, Institute of Food Safety, Bornsesteeg 45, NL-6708 PD Wageningen, The Netherlands

Hydrogenosomes are organelles that produce ATP and hydrogen¹, and are found in various unrelated eukaryotes, such as anaerobic flagellates, chytridiomycete fungi and ciliates². Although all of these organelles generate hydrogen, the hydrogenosomes from these organisms are structurally and metabolically quite different, just like mitochondria where large differences also exist³. These differences have led to a continuing debate about the evolutionary origin of hydrogenosomes^{4,5}. Here we show that the hydrogenosomes of the anaerobic ciliate *Nyctotherus ovalis*, which thrives in the hindgut of cockroaches, have retained a rudimentary genome encoding components of a mitochondrial electron transport chain. Phylogenetic analyses reveal that those proteins cluster with their homologues from aerobic ciliates. In addition, several nucleus-encoded components of the mitochondrial proteome, such as pyruvate dehydrogenase and complex II, were identified. The *N. ovalis* hydrogenosome is sensitive to inhibitors of mitochondrial complex I and produces succinate as a major metabolic end product—biochemical traits typical of anaerobic mitochondria³. The production of hydrogen, together with the presence of a genome encoding respiratory chain components, and biochemical

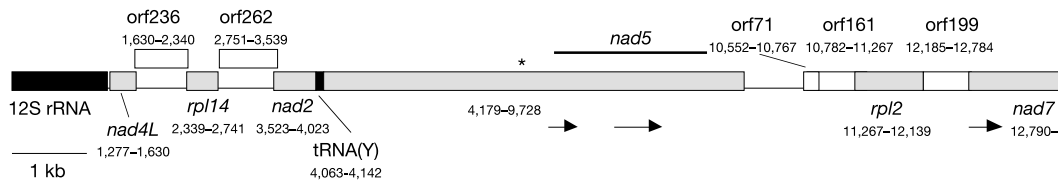


Figure 1 A 14,027-bp fragment (mtg 1) of the hydrogenosomal genome of *N. ovalis* var. *Blaberus* Amsterdam. Black boxes, RNA coding genes; shaded boxes, genes with significant similarity to mitochondrial genes; white boxes, unknown ORFs (named according to the number of codons); arrows, cDNAs identified so far. The numbers

indicate the nucleotide positions on the 14-kb clone (mtg 1). The longest ORF (4,179–9,728) contains a stretch with significant similarity to *nad5*. A potential start codon for a putative *nad5* transcript is marked with an asterisk.

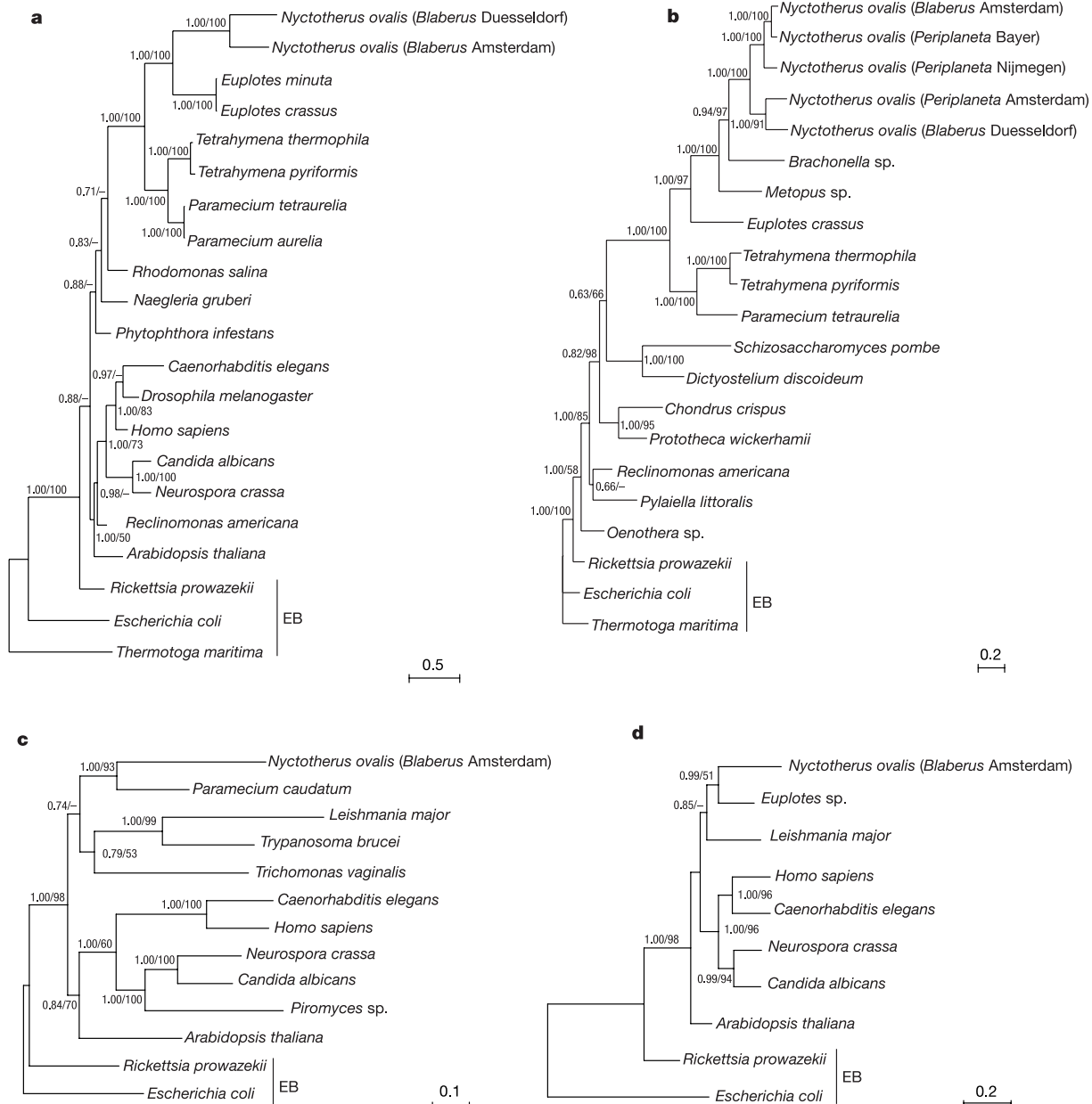


Figure 2 Phylogenetic analysis of hydrogenosomal genes. Both the organellar 12S (SSU) rRNA gene (**b**) and the nuclear hsp60 (**c**) reveal a ciliate ancestry for the hydrogenosome of *N. ovalis*. The same is true for the components of a 'mitochondrial' complex I, the *nad7* (49 kDa; organellar, **a**) and 51 kDa (nuclear, **d**) genes. The phylogenies were derived

using MrBayes and neighbour joining: the topologies correspond to the maximum-likelihood (MrBayes) approach, and the values at the nodes indicate the posterior probability for the partition and its bootstrap value, respectively. Only values higher than 50% are indicated. See Supplementary Information. EB, Eubacteria.

features characteristic of anaerobic mitochondria, identify the *N. ovalis* organelle as a missing link between mitochondria and hydrogenosomes.

Hydrogenosomes and their highly reduced relatives, mitosomes, generally lack an organelle genome^{5–8}, hampering clarification of their origin. Two models for the origin of hydrogenosomes are currently debated. The first posits that the ancestral mitochondrial endosymbiont gave rise to aerobically functioning mitochondria, which subsequently evolved into hydrogenosomes by the acquisition of genes encoding enzymes essential for an anaerobic metabolism^{9–13}. The second hypothesis presumes that hydrogenosomes and mitochondria originated from one and the same ancestral—facultatively anaerobic—(endo)symbiont, followed by specialization to aerobic and anaerobic niches during eukaryotic evolution^{13,14}.

To address this issue we investigated DNA in hydrogenosomes of *N. ovalis*, which was previously identified by immunocytochemical methods¹⁵. Intact *N. ovalis* hydrogenosomes isolated by cell fractionation contained DNA between 20 and 40 kilobases (kb) long. Long-range polymerase chain reaction (PCR) with this DNA with the use of specific primers for the hydrogenosomal small-subunit (SSU) ribosomal RNA¹⁵ and *nad7* (obtained earlier by PCR with degenerated primers) yielded a 12-kb fragment of the organellar genome. It encodes four genes of a mitochondrial complex I (*nad2*, *nad4L*, *nad5* and *nad7*), two genes encoding mitochondrial ribosomal proteins RPL 2 and RPL 14, and a ^{tyr}tRNA gene (Fig. 1). *Nad2* and *nad4L*, which are generally poorly conserved among ciliates, could be identified by using multiple sequence alignments and an analysis of their membrane-spanning domains as described

Table 1 *Nyctotherus ovalis* genes encoding mitochondrial proteins and RNAs

Type	Gene product	Localization of the gene	Codon use	cDNA	Target signal	Accession no.
Mt complex I	NAD2	H	Mt		No	AJ871267
	NAD4L	H	Mt		No	AJ871267
	NAD5	H	Mt	Yes*	No	AJ871267
	NAD7	H	Mt	Yes†	No	AJ871267
	24 kDa	N	Nuc		Yes	AY628688
	51 kDa	N	Nuc	Yes‡	Yes	AY608632
	75 kDa	N	Nuc		Yes	AJ871573
Mt complex II	SDH a	N	Nuc	Yes§	Yes	AY616152
	SDH b	N	Nuc	Yes	Yes	AY619980
Mt protein synthesis	Putative rRNA methyltransferase 2	N	Nuc			AJ871313
Mt ribosomal proteins	RPL 2	H	Mt			AJ871267
	RPL 14	H	Mt			AJ871267
	RPL 20	N	Nuc		?	AJ871314
Mt tRNA	tRNA tyrosine	H	rna			AJ871267
Mt rRNA	12S rRNA (SSU)	H	rna			AJ871267
Mt catabolism/energy metabolism	PDH E1α	N	Nuc	Yes¶	Yes	AY623917
	PDH E1β	N	Nuc	Yes**	Yes	AY628683
	PDH E2	N	Nuc	Yes††	Yes	AY623925
	[Fe] Hydrogenase	N	Nuc	Yes‡‡	Yes	AY608627
	Acetyl-coenzyme A synthase 2 (EC 6.2.1.1)	N	Nuc		No	AJ871315
	Adenylate kinase 2 (EC 2.7.4.3)	N	Nuc		No	AJ871316
	Adenylate kinase 2 isoform c = HK2418	N	Nuc			AJ871317
	Probable D-lactate dehydrogenase [cytochrome] (EC 1.1.2.4)	N	Nuc		Yes	AJ871318
	Succinyl-CoA ligase	N	Nuc		Yes	AJ871319
	Succinyl-CoA:3-ketoacid-coenzyme A transferase	N	Nuc		Yes	AJ871320
	Glycerol kinase	N	Nuc		?	AJ871321
	AAC	N	Nuc		No	AF480921
	Putative mt carrier protein PET8 YNL003C_Ch	N	Nuc		No	AJ871322
Mt import/processing	Mt processing peptidase alpha subunit	N	Nuc		?	AJ871323
	HSP 60	N	Nuc		Yes	AJ871324
	HSP 70	N	Nuc		No	AJ871325
	Heat shock protein HSP82 YMR186W_Ch	N	Nuc		No	AJ871326
	TOM 34	N	Nuc		No	AJ871327
	Mt protein import protein MAS5 (Protein YDJ1)	N	Nuc		?	AJ871328
	Stress-70 protein, mt precursor	N	Nuc		?	AJ871329

For a complete table see Supplementary Information. N, nucleus; H, hydrogenosome; nuc, nuclear; mt, mitochondrial; mtg 1, 12-kb clone of hydrogenosomal genome (AJ871267); ?, low-probability support so far because full-length cDNAs have not yet been isolated (the N terminus might be incomplete, or it might contain an in-frame intron or alternative start codons). Accession numbers for cDNAs: * AJ871574 and AJ871575; † AJ871576; ‡ AY608633 and AY608634; § AY616150 and AY616151; || AY619981; ¶ AY623919; ** AY628684; †† AY623926; ‡‡ AY608633 and AY608634.

Table 2 Glucose metabolism of *Nyctotherus ovalis*

Labelled end products	[U- ¹⁴ C]glucose		[6- ¹⁴ C]glucose	
	(μmol h ⁻¹ per mg protein)	(%)*	(μmol h ⁻¹ per mg protein)	(%)*
Acetate	427	53	467 ± 87	65 ± 22
Lactate	220	27	156 ± 116	20 ± 13
Succinate	112	14	79 ± 65	10 ± 7
Ethanol	44	5	29 ± 26	4 ± 3
CO ₂	205	—	ND	—
Formate	ND	—	ND	—

Cells were incubated for 48 h at 25 °C in micro-aerobic conditions in medium with either [U-¹⁴C]glucose or [6-¹⁴C]glucose. Excreted end products are shown as means ± s.d. of three independent experiments ([6-¹⁴C]glucose) or as the means of two independent experiments ([U-¹⁴C]glucose). Other excreted end products were less than 2% of the total excreted end products. ND, not detectable.

* Percentage of the total of acetate, lactate, succinate and ethanol.

previously¹⁶. Phylogenetic analysis revealed clustering of these genes with their homologues from the mitochondrial genomes of aerobic ciliates (Fig. 2, and Supplementary Information). All genes exhibit a characteristic mitochondrial codon-usage and lack amino-terminal extensions that could function as a mitochondrial targeting signal (Table 1). Complementary DNAs isolated for *nad5* and *nad7* show that they are transcribed. Translation with a nuclear genetic code from *N. ovalis*, rather than the ciliate mitochondrial code, leads to numerous stop codons (not shown). Five additional open reading frames (ORFs 236, 262, 71, 161 and 199) do not show significant sequence similarity to ORFs from the mitochondrial genomes accessible in the EMBL database. Two ORFs overlap with neighbouring ORFs as in other mitochondrial genomes¹⁷.

Macronuclear gene-sized chromosomes encoding the 24-kDa, 51-kDa and 75-kDa subunits of mitochondrial complex I and the Fp and Ip subunits of mitochondrial complex II were cloned with a PCR-based approach. These have a nuclear codon usage, are transcribed (Table 1), encode a putative N-terminal mitochondrial targeting signal and branch with their mitochondrial homologues from aerobic ciliates in phylogenetic analyses (Fig. 2, Table 1 and Supplementary Information). They are similar to the two complex I-like Ndh51 and Ndh24 proteins discovered in *Trichomonas vaginalis*^{18,19}, because a phylogenetic analysis including the mitochondrial homologues from *N. ovalis* and certain aerobic ciliates

reveals that all these proteins belong to a cluster of mitochondrial complex I homologues (see Supplementary Information). Thus, in *N. ovalis*, 7 of the 14 genes encoding core proteins of mitochondrial complex I, and two of the four proteins of mitochondrial complex II, have been identified so far. They are well conserved, are transcribed, and cluster with the mitochondrial homologues of their aerobic (ciliate) relatives, indicating that the hydrogenosomes of *N. ovalis* have retained parts of a functional mitochondrial electron-transport chain.

Hydrogenosomes of *N. ovalis* have typical mitochondrial cristae and contain cardiolipin¹¹. They are closely associated with endosymbiotic methanogens, which are biomarkers for hydrogen formation by the *N. ovalis* hydrogenosomes²⁰ (Fig. 3a). The organelles stain with Mitotracker Green FM and fluoresce with rhodamine 123, indicating the presence of a membrane potential (Fig. 3). Carbonyl cyanide *p*-trifluoromethoxyphenylhydrazone (FCCP) (5 μ M) prevented staining with rhodamine 123, indicating the possible presence of a proton gradient. Moreover, staining of the hydrogenosomes with rhodamine 123 was also prevented after incubation of the ciliates with rotenone, piericidin, fenazaquin and 1-methyl-4-phenylpyridinium (MPP⁺) (classical inhibitors of mitochondrial complex I (ref. 21)), but not with cyanide (1 mM) or antimycin A (inhibitors of mitochondrial complex III and IV; Fig. 3). Similarly, treatment with cyanide and salicylhydroxamic

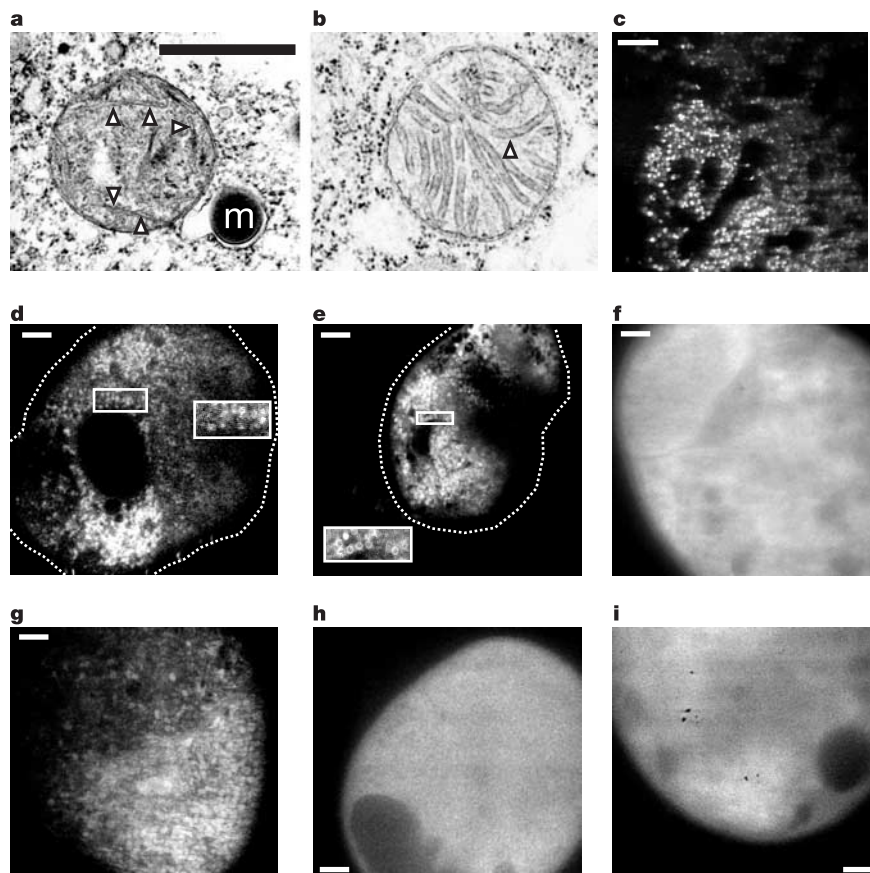


Figure 3 Hydrogenosomes of *N. ovalis* exhibit complex I activity. **a, b**, Electron micrographs of a hydrogenosome of *N. ovalis* (**a**) and a mitochondrion of *Euplotes* sp. (**b**). White arrowheads mark cristae; m, endosymbiotic methanogenic archaea^{15,20}. **c**, Fluorescence picture of *N. ovalis* hydrogenosomes (bright dots), which were released from the cell by gentle squashing after being stained *in vivo* with ethidium bromide. **d**, Rhodamine 123 (R123) also stains the hydrogenosomes, the only organelles matching the expected size (compare **a**, and Supplementary Information). **e**, Mitotracker green FM

stains the same organelles. The inserts in **d** and **e** (outside the ciliate) show the organelles seen in the box inside the ciliate. **f, h, i**, Incubation of living cells with inhibitors of mitochondrial complex I (MPP⁺ (**f**), piericidin (**h**) and rotenone (**i**))²¹ completely prevents staining of the organelles by R123. **g**, Incubation with cyanide (1 mM) or antimycin A (not shown) does not interfere with staining by R123. For additional information see the text and Supplementary Information. Scale bars, 1 μ m (**a**, also applies to **b**); 10 μ m (**c-i**).

acid (SHAM), inhibitors of mitochondrial complex IV of the respiratory chain and the plant-like alternative oxidase known from certain mitochondria³, respectively, neither killed *N. ovalis* nor interfered with its oxygen consumption under aerobic conditions (not shown). These observations not only indicate the absence of a functional complex III and IV and the absence of a terminal (plant-like) alternative oxidase, but also reveal the presence of a functional mitochondrial complex I as the source of the organellar proton gradient³. The oxygen consumption of *N. ovalis* observed under aerobic conditions is most probably a detoxification mechanism, and longer exposure to atmospheric oxygen kills the ciliates effectively.

Metabolic experiments using tracer amounts of uniformly labelled (U-¹⁴C)-glucose revealed that *N. ovalis* catabolizes glucose predominantly into acetate, lactate, succinate and smaller amounts of ethanol, in addition to CO₂ (Table 2). The presence of oxygen did not cause significant changes in the pattern of excreted end products (not shown). Notably, incubations in the presence of [6-¹⁴C]glucose did not result in the formation of labelled CO₂. Because ¹⁴C-labelled CO₂ is released from [6-¹⁴C]glucose by successive decarboxylations through multiple rounds in the Krebs cycle, the absence of labelled CO₂ after application of [6-¹⁴C]glucose indicates the absence of a complete Krebs cycle. The observed excretion of ¹⁴C-labelled CO₂ after incubation with [U-¹⁴C]glucose could be the result of either pyruvate dehydrogenase (PDH) activity, as in typical aerobic mitochondria, or pyruvate:ferredoxin oxidoreductase (PFO) activity, as in the hydrogenosomes of *T. vaginalis*. A third possibility for pyruvate catabolism, pyruvate formate lyase activity^{22,23}, can be excluded because no detectable amounts of formate were produced from [U-¹⁴C]glucose (Table 2). We failed to identify genes for PFO but succeeded in isolating three of the four PDH genes, namely the E1 α , E1 β and E2 subunits, which are expressed as cDNA, indicating that *N. ovalis* uses a mitochondrial PDH for oxidative decarboxylation. Significant amounts of ¹⁴C-labelled succinate from both [U-¹⁴C]glucose and [6-¹⁴C]glucose (Table 2) indicate that endogenously produced fumarate is used as a terminal electron acceptor, as in some anaerobic mitochondria³. Fumarate reduction in *N. ovalis* (to account for the production of succinate) is most probably catalysed by a membrane-bound complex II (see above; Table 1, and Supplementary Information), which is coupled to complex I through electron transport mediated by quinones³. Mass spectrometry coupled to liquid chromatography of lipid extracts from *N. ovalis* revealed the presence of small amounts of quinones (rhodoquinone 9 and menaquinone 8) at a concentration of about 1 pmol per mg protein (Supplementary Information). This amount is at least two orders of magnitude lower than in other eukaryotes known to possess anaerobic mitochondria producing succinate²⁴. The low concentration of quinones in *N. ovalis* cells might reflect the intermediate state of their hydrogenosomes, occupying a position between mitochondria (which contain a membrane-bound electron transport chain) and previously characterized hydrogenosomes (which do not)^{1,3–5,13,18}.

Although an F₀F₁-ATP synthase has not yet been identified, the hydrogenosome of *N. ovalis* has retained certain basal energy-generating functions of an aerobically functioning mitochondrion³. To explore the presence of additional 'mitochondrial' traits in *N. ovalis*, we performed a reciprocal Smith–Waterman sequence comparison between about 2,000 six-frame-translated clones from our genomic DNA library of *N. ovalis* and the yeast²⁵ and human²⁶ mitochondrial proteins. We identified 53 additional nuclear genes encoding potential mitochondrial proteins in addition to components of the mitochondrial import machinery (Table 1, and Supplementary Information).

In contrast, the hydrogenase of *N. ovalis* does not exhibit any mitochondrial traits. This hydrogenase is rather unusual in comparison with other eukaryotic hydrogenases because it seems to be a fusion of a [Fe] hydrogenase with two accessory subunits of

different evolutionary origin^{4,15}. These subunits should allow NADH reoxidation in combination with the [Fe] hydrogenase, because they exhibit a significant sequence similarity to the *hox F* and *hox U* subunits of β -proteobacterial [Ni–Fe] hydrogenases, in contrast to similar, recently described hydrogenosomal proteins (24 and 51 kDa) of putative mitochondrial origin from *T. vaginalis*^{18,19}. The 'mitochondrial' 24-kDa and 51-kDa genes of *N. ovalis* are clearly different from the above-mentioned hydrogenase modules and are likely to function in mitochondrial complex I (Fig. 3, Table 1, and Supplementary Information). Moreover, the catalytic centre of the hydrogenase (the H-cluster) clusters neither with any of the hydrogenases of *Trichomonas*, *Piromyces* and *Neocallimastix* studied so far, nor with any of the hydrogenase-related Nar proteins, which seem to be shared by all eukaryotes. Rather, the *N. ovalis* hydrogenase is more closely related to [Fe] hydrogenases from δ -proteobacteria^{4,15,27}. These observations suggest that the hydrogenase of *N. ovalis* has been acquired by lateral gene transfer.

It should be realized that the hydrogenosome of *N. ovalis* is so far unique and not representative of all hydrogenosomes, which seem to have evolved repeatedly and independently—albeit from the same ancestral mitochondrial-type organelle. All our data identify the hydrogenosome of *N. ovalis* as a ciliate-type mitochondrion that produces hydrogen. The presence of respiratory-chain activity of mitochondrial complex I and II, in combination with hydrogen formation, characterizes the *N. ovalis* hydrogenosome as a true missing link in the evolution of mitochondria and hydrogenosomes. □

Methods

Strains

Nyctotherus ovalis ciliates were isolated from the hindgut of the cockroach *Blaberus* sp. (strain Amsterdam), taking advantage of their unique (anodic) galvanotactic swimming behaviour²⁸. After the ciliate's arrival at the anode, cells were picked up with a micropipette, inspected individually under a dissecting microscope at $\times 40$ magnification, collected in an Eppendorf tube and washed three times with anaerobic electromigration buffer. Ciliates belonging to the *Euplotes crassus/vannus/minuta* complex were cultured in artificial sea water, feeding on bacteria growing on an immersed small piece of meat (approx. 1 g of beef steak). In addition, the ciliates were fed with *Escherichia coli*, which were added at weekly intervals. Ciliates were harvested by centrifugation.

Microscopy

Electron microscopy of *N. ovalis* and *Euplotes* sp. was performed as described previously^{11,15}.

Fluorescence microscopy was performed with a Noran OZ video-rate confocal microscope as described previously²⁹. Inhibitors were used in concentrations of 1 mM. They were dissolved in *N. ovalis* culture medium²⁸. The rotenone solution contained 10% dimethyl sulphoxide, the fenazaquin solution 1% dimethyl sulphoxide.

Metabolite and quinone determinations

Micro-aerobic incubations with *N. ovalis* were performed in rotating (20 r.p.m.) sealed incubation flasks containing 5 ml incubation medium (containing 10,000–15,000 cells). All incubations were performed for 48 h and contained either 10 μ Ci [U-¹⁴C]glucose or 10 μ Ci [6-¹⁴C]glucose (2.07 GBq mmol⁻¹), both from Amersham. Incubations were terminated by the addition of 300 μ l 6 M HCl to lower the pH from 7.2 to 2.0. *N. ovalis* cells were separated from the medium by centrifugation (5 min at 500g and 4 °C); excreted end products were analysed by anion-exchange chromatography on a Dowex 1X8 column. Quinones were separated by liquid chromatography and detected with a Sciex API 300 triple quadrupole mass spectrometer (see Supplementary Information).

Isolation of organellar DNA

N. ovalis cells were washed once with isolation buffer (0.35 M sucrose, 10 mM Tris-HCl pH 7, 2 mM EDTA) and disrupted in a Dounce homogenizer. Nuclei were centrifuged at 3,000g for 5 min, and organelles were pelleted from the supernatant at 10,000g for 30 min. Genomic DNA was isolated by using standard procedures or after lysis of the cells in 8 M guanidinium chloride.

Genomic DNA library

Gene-sized chromosomes were randomly amplified by PCR with telomere-specific primers, size-fractionated in agarose gels, and cloned in pGEM-T-easy (Promega). Clones with sizes between 0.5 and 5 kb were end-sequenced and analysed manually by TBLASTX (<http://www.ncbi.nlm.nih.gov/BLAST>). Searches were conducted with BLASTN and FASTA.

c-DNA library

RNA was isolated with the RNeasy Plant minikit (Qiagen). cDNA was prepared with the

Smart Race cDNA amplification kit (Clontech). Expressed sequence tags were amplified by PCR with the universal adapter primer provided with the kit and the various, specific internal primers.

Complete macronuclear gene-sized chromosomes

Telomere-specific primers in combination with internal gene sequences allow a straightforward recovery of the complete gene³⁰. The specific (internal) primers were based on the DNA sequences of internal fragments of the various genes, which were recovered previously by PCR with degenerated primers for conserved parts of the various genes.

Phylogenetic analysis

Protein sequences were aligned with ClustalW and Muscle; unequivocally aligned positions were selected with Gblocks or manually. Phylogenies were inferred with maximum likelihood by using a discrete gamma-distribution model with four rate categories plus invariant positions and the Poisson amino acid similarity matrix, and neighbour joining as implemented in ClustalW, correcting for multiple substitutions with the Gonnert amino acids identity matrix, and bootstrapping with 100 samples.

ORFs with a lower size limit of 100 nucleotides were identified with ORF Finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). tRNAs were identified with tRNAscan-SE (<http://www.genetics.wustl.edu/eddy/tRNAscan-SE>). Potential mitochondrial import signals were detected with MITOP (<http://mips.gsf.de/cgi-bin/proj/medgen/mitofilter>). Sequence searches were performed with BLASTX (<http://www.ncbi.nlm.nih.gov/BLAST>), BLASTN and FASTA. For references on phylogenetic analysis see Supplementary Information.

Received 18 August 2004; accepted 7 January 2005; doi:10.1038/nature03343.

- Müller, M. The hydrogenosome. *J. Gen. Microbiol.* **39**, 2879–2889 (1993).
- Roger, A. J. Reconstructing early events in eukaryotic evolution. *Am. Nat.* **154**, S146–S163 (1999).
- Tielens, A. G. M., Rotte, C., van Hellemond, J. J. & Martin, W. Mitochondria as we don't know them. *Trends Biochem. Sci.* **27**, 564–572 (2002).
- Embley, T. M. *et al.* Hydrogenosomes, mitochondria and early eukaryotic evolution. *IUBMB Life* **55**, 387–395 (2003).
- Dyall, S. D., Brown, M. T. & Johnson, P. J. Ancient invasions: From endosymbionts to organelles. *Science* **304**, 253–257 (2004).
- van der Giezen, M., Sjöllerna, K. A., Artz, R. R., Alkema, W. & Prins, R. A. Hydrogenosomes in the anaerobic fungus *Neocallimastix frontalis* have a double membrane but lack an associated organelle genome. *FEBS Lett.* **408**, 147–150 (1997).
- Clemens, D. L. & Johnson, P. J. Failure to detect DNA in hydrogenosomes of *Trichomonas vaginalis* by nick translation and immunomicroscopy. *Mol. Biochem. Parasitol.* **106**, 307–313 (2000).
- Leon-Avila, G. & Tovar, J. Mitosomes of *Entamoeba histolytica* are abundant mitochondrion-related remnant organelles that lack a detectable organelle genome. *Microbiol.* **150**, 1245–1250 (2004).
- Fenchel, T. & Finlay, B. J. *Ecology and Evolution in Anoxic Worlds* (Oxford University Press, Oxford, UK, 1995).
- Embley, T. M., Horner, D. A. & Hirt, R. P. Anaerobic eukaryote evolution: hydrogenosomes as biochemically modified mitochondria? *Trends Ecol. Evol.* **12**, 437–441 (1997).
- Voncken, F. *et al.* Multiple origins of hydrogenosomes: functional and phylogenetic evidence from the ADP/ATP carrier of the anaerobic chytrid *Neocallimastix* sp. *Mol. Microbiol.* **44**, 1441–1454 (2002).
- van der Giezen, M. *et al.* Conserved properties of hydrogenosomal and mitochondrial ADP/ATP carriers: a common origin for both organelles. *EMBO J.* **21**, 572–579 (2002).
- Martin, W., Hoffmeister, M., Rotte, C. & Henze, K. An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol. Chem.* **382**, 1521–1539 (2001).
- Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
- Akhmanova, A. *et al.* A hydrogenosome with a genome. *Nature* **396**, 527–528 (1998).
- Brunk, C. F., Lee, L. C., Tran, A. B. & Li, J. Complete sequence of the mitochondrial genome of *Tetrahymena thermophila* and comparative methods for identifying highly divergent genes. *Nucleic Acids Res.* **31**, 1673–1682 (2003).
- Burger, G., Gray, M. W. & Lang, B. F. Mitochondrial genomes: anything goes. *Trends Genet.* **19**, 709–716 (2003).
- Dyall, S. D. *et al.* Non-mitochondrial complex I proteins in a hydrogenosomal oxidoreductase complex. *Nature* **431**, 1103–1107 (2004).
- Hrdy, I. *et al.* *Trichomonas* hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. *Nature* **432**, 618–622 (2004).
- van Hoek, A. H. A. M. *et al.* Multiple acquisition of methanogenic archaeal symbionts by anaerobic ciliates. *Mol. Biol. Evol.* **17**, 251–258 (2000).
- Degli Esposti, M. Inhibitors of NADH-ubiquinone reductase: an overview. *Biochim. Biophys. Acta* **1364**, 222–235 (1998).
- Akhmanova, A. *et al.* A hydrogenosome with pyruvate formate-lyase: anaerobic chytrid fungi use an alternative route for pyruvate catabolism. *Mol. Microbiol.* **32**, 1103–1114 (1999).
- Boxma, B. *et al.* The anaerobic chytridiomycete fungus *Piromyces* sp. E2 produces ethanol via pyruvate:formate lyase and an alcohol dehydrogenase E. *Mol. Microbiol.* **51**, 1389–1399 (2004).
- van Hellemond, J. J., Klockiewicz, M., Gaasenbeek, C. P. H., Roos, M. H. & Tielens, A. G. M. Rhodoquinone and complex II of the electron transport chain in anaerobically functioning eukaryotes. *J. Biol. Chem.* **270**, 31065–31070 (1995).
- Sickmann, A. *et al.* The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc. Natl Acad. Sci. USA* **100**, 13207–13212 (2003).
- Cotter, D., Guda, P., Fahy, E. & Subramaniam, S. MitoProteome: mitochondrial protein sequence database and annotation system. *Nucleic Acids Res.* **32**, D463–D467 (2004).
- Voncken, F. G. J. *et al.* A hydrogenosomal [Fe]-hydrogenase from the anaerobic chytrid *Neocallimastix* sp. L2. *Gene* **284**, 103–112 (2002).
- van Hoek, A. H. A. M. *et al.* Voltage-dependent reversal of anodic galvanotaxis in *Nyctotherus ovalis*. *J. Eukaryotic Microbiol.* **46**, 427–433 (1999).

- Koopman, W. J. H. *et al.* Membrane-initiated Ca^{2+} signals are reshaped during propagation to subcellular regions. *Biophys. J.* **81**, 57–65 (2001).
- Curtis, E. A. & Landweber, L. F. Evolution of gene scrambling in ciliate micronuclear genes. *Ann. NY Acad. Sci.* **870**, 349–350 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L. Landweber, J. Wong and W.-J. Chang for advice on the cloning of complete minichromosomes and for sharing the first sequence of a PDH gene in *N. ovalis*; S. van Weelden and H. de Rook for help in the metabolic studies; J. Brouwers for analysis of the quinones; G. Cremers, L. de Brouwer, A. Ederveen, A. Grootemaat, M. Hachmang, S. Huver, S. Jannink, N. Jansse, R. Janssen, M. Kwantes, B. Penders, G. Schilders, R. Talens, D. van Maassen, H. van Zoggel, M. Veugeling and P. Wijnhoven for help with the isolation of various *N. ovalis* sequences; and K. Sjöllerna for electron microscopy. G.W.M.v.d.S., S.Y.M.-v.d.S. and G.R. were supported by the European Union 5th framework grant 'CIMES'. This work was also supported by equipment grants from ZON (Netherlands Organisation for Health Research and Development), NWO (Netherlands Organisation for Scientific Research), and the European Union 6th framework programme for research, priority 1 "Life sciences, genomics and biotechnology for health" to W.J.H.K..

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H.P.H. (j.hackstein@science.ru.nl). Sequences have been deposited at the EMBL database under accession numbers AF480921, AJ871267, AJ871313–AJ871361, AJ871573–AJ871576, AY608627, AY608632–AY608634, AY616150–AY616152, AY619980, AY619981, AY623917, AY623919, AY623925, AY623926, AY628683, AY628684, AY628688.

Image segmentation and lightness perception

Barton L. Anderson¹ & Jonathan Winawer²

¹University of New South Wales, School of Psychology, Sydney, New South Wales 2052, Australia

²Massachusetts Institute of Technology, Brain and Cognitive Sciences, Cambridge, Massachusetts 02139, USA

The perception of surface albedo (lightness) is one of the most basic aspects of visual awareness. It is well known that the apparent lightness of a target depends on the context in which it is embedded^{1–6}, but there is extensive debate about the computations and representations underlying perceived lightness. One view asserts that the visual system explicitly separates surface reflectance from the prevailing illumination and atmospheric conditions in which it is embedded^{7–10}, generating layered image representations. Some recent theory has challenged this view and asserted that the human visual system derives surface lightness without explicitly segmenting images into multiple layers^{11,12}. Here we present new lightness illusions—the largest reported to date—that unequivocally demonstrate the effect that layered image representations can have in lightness perception. We show that the computations that underlie the decomposition of luminance into multiple layers under conditions of transparency can induce dramatic lightness illusions, causing identical texture patches to appear either black or white. These results indicate that mechanisms involved in decomposing images into layered representations can play a decisive role in the perception of surface lightness.

The amount of light projected to the eyes (luminance) is determined by a number of factors: the illumination that strikes visible surfaces, the proportion of light reflected from the surface and the amount of light absorbed, reflected or deflected by the prevailing atmospheric conditions (such as haze or other partially transparent media). Only one of these factors, the proportion of light reflected (lightness), is associated with an intrinsic property of

letters to nature

surfaces, and hence is of special interest to the visual system. To accurately recover lightness, the visual system must somehow disentangle the contributions of surface reflectance from the illumination and atmospheric conditions in which it is embedded. One theoretical view asserts that the visual system explicitly decomposes images into a set of separate maps or layers, corresponding to the separate physical contributions to retinal luminance^{9,10}. However, there is a growing body of data showing that the visual system can make systematic errors in estimating surface reflectance¹¹, the opacity of transparent surfaces or media¹³ and the amount of illumination striking a surface¹⁴. These errors have led some to question whether the visual system explicitly decomposes images

into their constituent physical sources, and to suggest that the visual system uses computational 'short cuts' to generate representations of surface lightness. Such theories have suggested that the visual system divides an image into two-dimensional regions (or 'frameworks') rather than layers. In such models, lightness is derived through processes that bias the highest luminance to appear white¹¹, and/or by using statistical estimation techniques within local image regions to compute reflectance¹²; no explicit decomposition of the image into separate layers occurs.

One of the most widely used techniques to explore context effects in lightness perception is to embed identical target patches in different surrounds. Most studies with this method have used

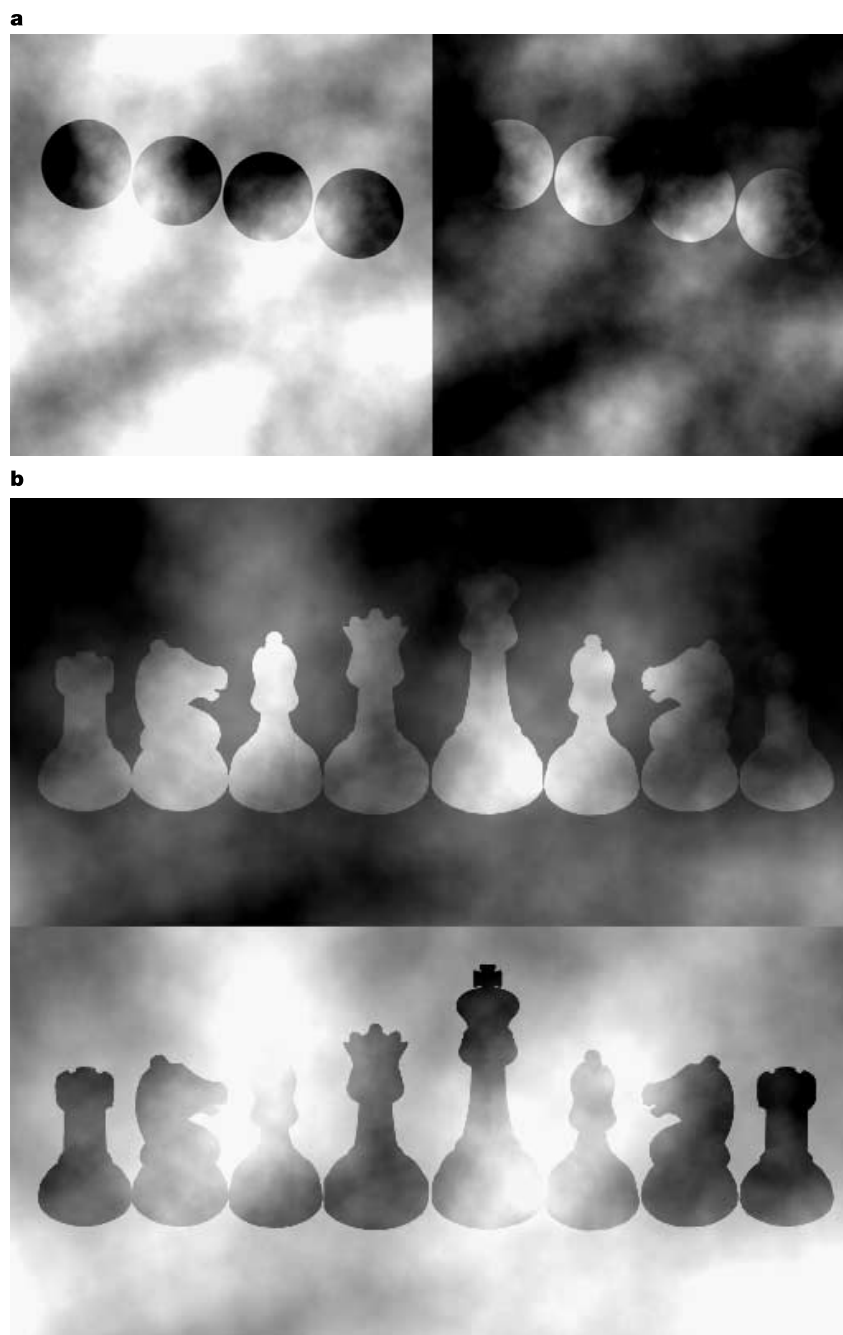


Figure 1 Static versions of the lightness illusions studied in our experiment (see also Supplementary Video 1). In **a**, the corresponding textured disks on the dark and light surrounds are physically identical, and in **b** the corresponding chess pieces on the two

surrounds are identical. In both cases, the figures on the dark surround appear as light objects visible through dark haze, whereas the figures on the light surround appear as dark objects visible through light haze.

untextured patches containing a uniform reflectance or luminance. In such images, it is usually difficult if not impossible to determine whether a target is decomposed into multiple sources; the target region simply appears to be a particular shade of grey. To assess the role of layered image representations in perceived lightness, we devised a new set of stimuli that would make a layered decomposition perceptually apparent if it was occurring (Fig. 1). We generated textured images that contain a continuous distribution of luminance values and we also manipulated geometric and luminance relationships known to play a role in the segmentation of surfaces into multiple layers^{13,15–17}. A common ‘seed’ texture was used to create both the targets and the surrounds (see Methods). The target regions in the two images were identical; only the surrounds differed. One surround was made lighter than the seed image, and the other darker. The target regions were placed in the same relative position (compared to the seed image) on each of the two surrounds. The critical image properties manipulated using this technique were the polarity and magnitude of contrast between the textures and their surrounds. In the image with the dark surround, the polarity of the surround–target border was dark–light along its entire length of the border (respectively); in the light surround, the surround–target border was light–dark. Contrast magnitude varied continuously over both surround–target borders. As can be seen in Fig. 1 (and even more dramatically in the moving versions in Supplementary Video 1), this manipulation caused a striking difference in appearance between the central targets. For the dark surround, the target regions appeared white, visible through dark, partially transparent clouds; for the light surround, the identical targets appeared black, visible through light clouds. Note that the fluctuations in contrast magnitude along the target–surround border appear as variations in the opacity of the transparent layer; this is in keeping with recent research demonstrating that the visual system uses variations in contrast magnitude to compute the opacity of transparent layers¹³.

We performed a lightness matching experiment to determine what was responsible for the perceived lightness of the targets in these images. The targets in Fig. 1 contain luminance values that span the range from white to black. One explanation of the perceived lightness difference in these images is that the two surrounds cause the target regions to be decomposed in two very different ways. For the targets on the light surround, the darkest pixels appear to form an unobscured view of the distant surface, and lighter pixels appear to be a combination of a light transparent layer and a dark distant surface. For the targets on the dark surround, the lightest pixels appear to form an unobscured view of the distant surface, and darker pixels appear to be a combination of a dark transparent layer and a light distant surface. This decomposition is consistent with recent theory that asserts that the visual system makes use of the sign and magnitude of image contrast to determine those portions of surfaces that are in plain view and those surface regions that are obscured by transparent media¹⁵. In this account, the highest contrast regions are seen in plain view (the lightest and darkest pixels on the dark and light surrounds respectively), and lower contrast values are seen through a contrast-reducing medium (where the opacity of the transparent layer is proportional to the amount by which the highest contrast is reduced). According to this view, the entire target regions in each figure are seen to have a single lightness value as determined by the pixels in plain view. If this analysis is correct, then the perceived lightness of the light target should correspond to the perceived lightness of the brightest pixels in the target region, and the matches for the dark target should correspond to the perceived lightness of the darkest pixels in the target region.

To test this hypothesis, we varied the range of intensities in the target region (that is, its contrast), and observers adjusted the luminance of a test patch until it appeared to match the lightness of the target patches. A control experiment was performed to

determine the contribution of simple contrast enhancement mechanisms of the surrounds on uniform grey patches (see Methods and Supplementary Video 2). Results of these experiments are shown in Fig. 2. The solid lines depict the lightest and darkest pixels in the target patches (Fig. 2a). Lightness matches by the observers correspond closely to these lines, but there is also a slight overestimation of the lightness of the targets on both surrounds (the apparent saturation of the matches to the light target reflects the limited luminance range of the monitors; the luminance setting in these regions is simply the maximal available). This overestimation plays only a small part in the magnitude of the lightness transformation reported here, but it is consistent with data showing that the visual system normalizes luminance in a manner that generates a bias for observers to perceive the highest luminance as white¹¹. The contrast control experiment (Fig. 2b) showed that simple contrast enhancement processes produce a much smaller illusion (only 11% as large as the largest lightness difference with the textured targets), and therefore cannot account for the illusions in Fig. 1.

These results are consistent with the hypothesis that the transformation in lightness observed in Fig. 1 arises from segmentation processes involved in the perception of transparency. If this is correct, then such lightness transformations should be abolished if the conditions critical for inducing the perception of transparency

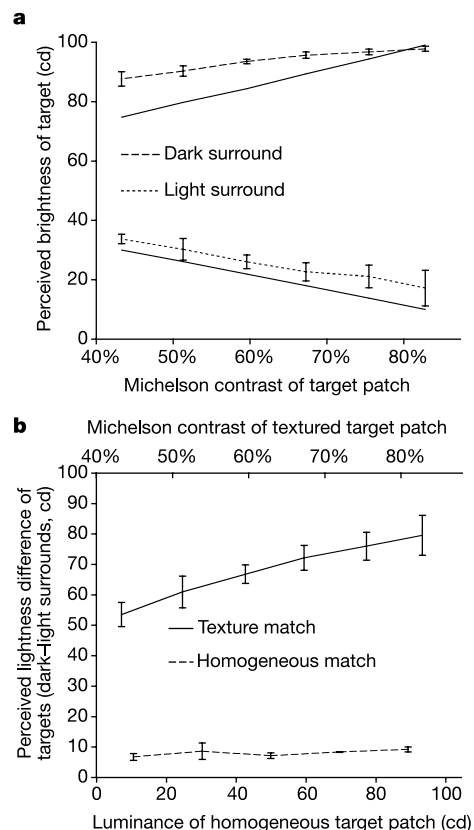


Figure 2 Lightness matching data. **a**, Data obtained using the moving version of the illusion. The light and dark surrounds were held constant, and the contrast (luminance range) of the circular target patches were varied. Solid lines depict the luminance of the lightest (upper line) and darkest (lower line) pixels in the target patch. The data are close to these lines, but there is a bias for observers to report the targets on both light and dark surrounds as lighter than these values. **b**, The data in **a** are compared with the control experiment using homogeneous targets that varied in luminance (plotted as a difference between the matches to the targets on the dark and light surrounds, respectively). The contrast effects are only 11% as large as the largest effects with the textured targets. Error bars depict the standard error of the mean for three subjects.

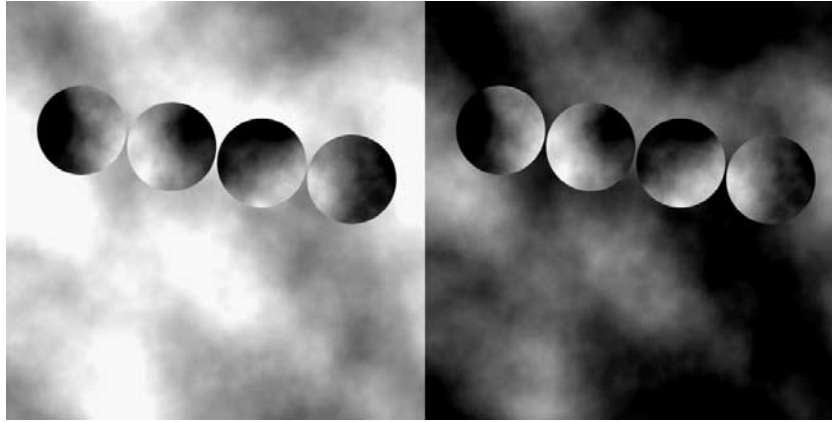


Figure 3 Transparency control experiment. The same targets and surrounds are used as in Fig. 1a, except that the surrounds have been rotated by 90° (see also Supplementary

Video 3). This rotation destroys both the geometric and luminance conditions needed to evoke a percept of transparency, and also destroys the lightness illusion.

are removed but all other aspects of the display remain unchanged. To test this hypothesis, the light and dark surrounds were simply rotated by 90°, destroying both the geometric conditions (the continuity of the textures in the targets and the surround) and the luminance relationships needed to induce the perception of transparency (the contrast polarity now reverses along the borders of both set of targets). As can be seen in Fig. 3 (and in Supplementary Video 3), this manipulation destroys the lightness difference observed in Fig. 1, demonstrating that transparency plays a critical role in the lightness transformations that occur in these displays.

The phenomena reported here provide new insights into the computations underlying lightness perception. The data are consistent with the view that lightness perception cannot be understood with low-level mechanisms such as lateral inhibition, as such mechanisms produce a much smaller illusion^{2,5,11}. Recent lightness models that omit computations that generate layered image representations also fail to account for the phenomena reported here. Such models decompose images into a set of discrete two-dimensional sub-regions, and estimate lightness within each sub-region separately using principles of anchoring¹¹ or statistical estimation¹². Note, however, that the perceived transmittance of the transparent layer appears to vary continuously over the entire image in Fig. 1. It is unclear how such models could account for these phenomena.

It should be noted that layered image representations underlying the illusions reported here are conceptually related to figure-ground reversals¹⁶. Note, however, that traditional figure-ground reversals involve shifts between image regions that occupy different regions of space, whereas the phenomena reported here involve image regions along the same visual directions, and hence involve the same set of pixels. To see this, consider the segmentation processes involved in viewing a surface through an occluding mesh or screen. In such contexts, the visual system must determine which image regions correspond to the occluding screen and which regions correspond to the underlying surface visible through the holes in the screen. This conception of transparency is readily generalized to continuous media by simply allowing the holes to become infinitesimally small. If the perceived depth order of the surfaces is reversed, then the perceived lightness of the two layers will shift as well, as can be experienced in Fig. 1.

There is a growing body of data demonstrating that a variety of factors influence perceived lightness, including surface curvature⁴, surface orientation¹⁸, depth^{2,10,11,17} and simply the number of different surfaces in a scene¹¹. Previous research has shown that transformations in perceived lightness can occur in images that

induce percepts of transparency in stereoscopic displays¹⁷. However, the causal role of such segmentation processes in lightness perception has not been previously established for monocular images. The data presented here provide unequivocal evidence that segmentation processes underlying the formation of layered image representations can play a critical and dramatic role in lightness perception. Theories of lightness perception that do not include such processes are at best incomplete. □

Methods

Textures

Textured 'seed' images were generated in Matlab as grey-scale noise with a specified power spectrum that varied as $(1/f^4)$, 512×512 pixels. The different frequency components were summed with random phases and orientations. The target and background images were spatially identical to the seed image, but differed in the range of luminance values. The target image had 99% Michelson contrast, with intensities ranging from 1 to 96 cd. For the surrounds, the luminance ranges were compressed and either shifted up or down. For the light surround (which gives rise to the percept of a dark target seen through light clouds), the range was 36 to 96 cd (45% contrast) and for the dark surround, the range was 1 to 77 cd (98% contrast). The illusions (static, Fig. 1; moving, Supplementary Video 1) were created by aligning the target texture with one of the surround textures and then showing the target through a circular aperture on either the light or dark surround. The multiple apertures in Fig. 2 represent the effect of motion. For the control demonstrations (Fig. 3), the identical targets were used but the surrounds were rotated by 90°. This caused the polarity relationships between the target patch and the surround to vary, destroying the percept of transparency and the lightness illusion.

Matching experiment

To quantify the perceived lightness in Fig. 1, subjects adjusted a test patch to match the perceived lightness of the targets. Using the stimuli described above, observers were presented with a circular target (3° in diameter) moving back and forth horizontally (one cycle every 5 s) on either a light or a dark square surround (15° per side). They adjusted the luminance of a uniform, square test patch (2°) on a black and white checkered background (3°) until the test patch appeared to be the same lightness as the moving target. Subjects had unlimited time to make the matches. Stimuli consisted of either the light or dark surrounds depicted in Fig. 1 along with one of six target stimuli similar to the target in Fig. 1, but ranging in contrast from 0.43 to 0.82. To measure the contribution of simple contrast enhancement processes to the illusion, homogenous grey disks that were identical in size to the textured targets were also presented (ranging in luminance from 10 to 89 cd), and the same matching task was used. All combinations of backgrounds (one light and one dark) and targets (six patterned and five grey) were presented five times each in random order to each subject, for a total of 55 trials.

Stimuli were generated using Vision Shell software and were presented on an Apple Macintosh G4 computer using a Lacie (electron22blue) monitor that was calibrated and linearized before testing observers. Viewing distance was 57 cm. One subject, CB, was naïve and the other two observers were the authors.

Received 13 October; accepted 13 December 2004; doi:10.1038/nature03271.

1. Koffka, K. *Principles of Gestalt Psychology* (Harcourt, Brace and World, New York, 1935).
2. Gilchrist, A. L. Perceived lightness depends on spatial arrangement. *Science* **195**, 185–187 (1977).
3. Gilchrist, A. L. When does perceived lightness depend on perceived spatial arrangement? *Percept. Psychophys.* **28**, 527–538 (1980).

4. Knill, D. C. & Kersten, D. Apparent surface curvature affects lightness perception. *Nature* **351**, 228–230 (1991).
5. Adelson, E. H. Perceptual organization and the judgment of brightness. *Science* **262**, 2042–2044 (1993).
6. Eagleman, D. M., Jacobson, J. E. & Sejnowski, T. J. Perceived luminance depends on temporal context. *Nature* **428**, 854–856 (2004).
7. Bergstrom, S. S. Common and relative components of reflected light as information about the illumination, colour, and three-dimensional form of objects. *Scand. J. Psychol.* **18**, 180–186 (1977).
8. Gilchrist, A. L. The perception of surface blacks and whites. *Sci. Am.* **240**, 112–123 (1979).
9. Barrow, H. G. & Tenenbaum, J. in *Computer Vision Systems* (eds Hanson, A. R. & Riseman, E. M.) 3–26 (Academic, New York, 1978).
10. Anderson, B. L. A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions. *Perception* **26**, 419–453 (1997).
11. Gilchrist, A. et al. An anchoring theory of lightness perception. *Psychol. Rev.* **106**, 795–834 (1999).
12. Adelson, E. H. in *The New Cognitive Neurosciences* 2nd edn (ed. Gazzaniga, M.) 339–351 (MIT Press, Cambridge, Massachusetts, 1999).
13. Singh, M. & Anderson, B. L. Toward a perceptual theory of transparency. *Psychol. Rev.* **109**, 492–519 (2002).
14. Rutherford, M. D. & Brainard, D. H. Lightness constancy: a direct test of the illumination estimation hypothesis. *Psychol. Sci.* **13**, 142–149 (2002).
15. Anderson, B. L. The role of occlusion in the perception of depth, lightness, and opacity. *Psychol. Rev.* **110**, 762–784 (2003).
16. Anderson, B. L. The role of perceptual organization in White's illusion. *Perception* **32**, 269–284 (2003).
17. Anderson, B. L. Stereoscopic surface perception. *Neuron* **24**, 919–928 (1999).
18. Boyaci, H., Maloney, L. T. & Hersh, S. The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *J. Vis.* **3**, 541–553 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank N. Witthoft for suggesting the chessboard variant of the illusion and C.U. Jo for inspiration and support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.L.A. (bart.a@unsw.edu.au).

Cross-presentation by intercellular peptide transfer through gap junctions

Joost Neijssen^{1*}, Carla Herberths^{1*}, Jan Wouter Drijfhout², Eric Reits¹, Lennert Janssen¹ & Jacques Neeffjes¹

¹Division of Tumor Biology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

²Department of Immunohematology and Blood Transfusion Leiden, University Medical Center, Albinusdreef 2, 2333RC Leiden, The Netherlands

* These authors contributed equally to this work

Major histocompatibility complex (MHC) class I molecules present peptides that are derived from endogenous proteins¹. These antigens can also be transferred to professional antigen-presenting cells in a process called cross-presentation, which precedes initiation of a proper T-cell response^{2,3}; but exactly how they do this is unclear. We tested whether peptides can be transferred directly from the cytoplasm of one cell into the cytoplasm of its neighbour through gap junctions. Here we show that peptides with a relative molecular mass of up to ~1,800 diffuse intercellularly through gap junctions unless a three-dimensional structure is imposed. This intercellular peptide transfer causes cytotoxic T-cell recognition of adjacent, innocent bystander cells as well as activated monocytes. Gap-junction-mediated peptide transfer is restricted to a few coupling cells owing to the high cytosolic peptidase activity. We present a mechanism of antigen acquisition for cross-presentation that couples the antigen presentation system of two adjacent cells and

is lost in most tumours: gap-junction-mediated intercellular peptide coupling for presentation by bystander MHC class I molecules and transfer to professional antigen presenting cells for cross-priming.

MHC class I molecules present peptides to the immune system for surveillance by CD8⁺ cytotoxic T cells (CTL). Because intracellular antigens and antigenic peptides usually cannot traverse membranes, only endogenous peptides can be presented by MHC class I molecules¹. Antigenic peptides from infected cells are thus exclusively loaded on the cell's own MHC class I molecules and not on those of innocent bystander cells. This concept has been challenged by a process called cross-presentation^{2,3}. Cross-presentation implies the transfer of antigenic (usually intracellular) antigens from diseased cells to professional antigen-presenting cells (APC) such as dendritic cells, activated monocytes or Langerhans cells^{2–4}. The APCs subsequently present these antigenic peptides on their own MHC class I molecules, and migrate to draining lymph nodes where activation and expansion of the specific CD8⁺ T-cell population occurs. Cross-presentation requires that antigens somehow enter the MHC class I presentation pathway of an APC.

In this study, we investigated the possibility of direct gap-junction-mediated transfer of antigens between the cytoplasm of two adjacent cells. Gap junctions are assemblies of intercellular channels that form an integral part of multicellular organisms. A functional channel is formed when a hemichannel, composed of six connexin molecules, assembles with a hemichannel from an adjacent cell⁵. The resulting gap junctions electrically couple cells by direct exchange of ions and allow exchange of nutrients and second messengers. Gap junctions are thought to be non-specific channels that allow passive diffusion of molecules with a relative molecular mass of up to 1,000 (M_r 1K)⁶ and intracellular signalling controls the gating⁷. Connexin 43 (Cx43) is broadly expressed, whereas the other connexin family members are expressed in specific tissues only. Cx43 is also expressed in various haematopoietic cells like follicular dendritic cells, B cells, activated lymphocytes and monocytes⁸. Importantly, many tumour cells are uncoupled from their environment, for example after inactivation of their gap junctions by ras, src and neu oncogenes or by APC deficiency^{9,10}. Viral proteins of the herpesvirus HSV-2 (ref. 11) and the human papilloma virus HPV-16 (ref. 12) are able to close gap junctions of infected cells. In addition, gap junction intercellular communication seems to be important for the bystander effect in cancer gene therapy¹³.

To visualize peptide transfer between cells, we used A431 cells. This human squamous carcinoma cell line does not express gap junctions, as shown by biochemical and biophysical techniques¹⁴. A431 cells were stably transfected with human Cx43 (Fig. 1a), resulting in functional gap junctions. To study peptide transfer between cells, stable fluorescently labelled (FL-) peptides were synthesized. These peptides are not degraded in cells because they are composed of D-amino acids with a protective group at the amino terminus¹⁵. A 9-mer FL-peptide was introduced in A431/Cx43 and control A431 cells by micro-injection, together with dextran-TexasRed (TxR) (M_r 70K) as an injection marker. Cells were subsequently analysed by confocal laser scanning microscopy (CLSM) (Fig. 1b) and transfer was quantified in both cell lines (Fig. 1c). Whereas dextran-TxR is maintained in the micro-injected cell, the 9-mer peptide diffused into surrounding cells only when the cells expressed Cx43. Closure of gap junctions by chemical inhibitors such as 2-APB (ref. 16) prevented this intercellular peptide transfer between A431/Cx43 cells.

To test the efficiency of gap-junction-mediated peptide transfer, small groups of A431/Cx43 cells were grown on coverslips. Peptides of various lengths were micro-injected along with dextran-TxR and the rate of transfer was determined using fluorescence recovery after photobleaching (FRAP) techniques. Peptides were permitted to

migrate to neighbouring cells before the start of a FRAP experiment. Subsequently, peptide fluorescence was ablated by photobleaching one neighbouring cell and the fluorescence recovery measured (Fig. 2a). Reappearance of FL-peptides was readily observed with this protocol as illustrated in Fig. 2b, due to peptide exchange between the bleached and surrounding cells. The half-time of maximal FL-peptide recovery ($t_{1/2}$) was determined for FL-peptides varying from four up to ten amino acids (M_r 1,082–1,851) in length in A431/Cx43 cells (Fig. 2c). Immunologically relevant peptides, normally between eight to ten amino acids in length, can be transferred through Cx43 gap junctions from the cytosol of one cell into its neighbour. This implies that intercellular peptide spreading can occur, possibly resulting in the elimination of innocent bystander cells when connected by gap junctions to the infected cell.

The size of a peptide seems to determine the rate of transfer between cells, because this rate decreases for longer peptides (Fig. 2c). It can be envisaged that longer peptides inherently contain more secondary structure, hampering gap-junctional transfer. To test whether structure or molecular mass are determining factors, an 8-mer peptide was synthesized in two conformational states—linear (probably flexible) and circular—and the efficiency of Cx43-mediated peptide transfer was determined in A431/Cx43 cells (Fig. 2d). Only the 'linear' peptide was efficiently transferred between cells, indicating that its three-dimensional structure is an important factor for gap-junction-mediated peptide transfer. Probably gap-junctional transfer is restricted to peptides that are able to adopt an extended conformation.

Although stable peptides can be transferred intercellularly, the high cytosolic peptidase activity might restrict or even prevent intercellular transfer of normal peptides *in vivo*¹⁵. We previously

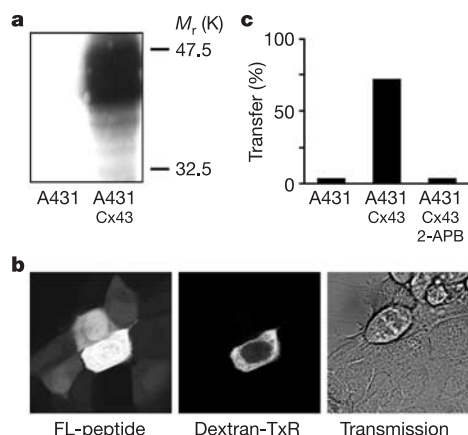


Figure 1 Gap-junction-dependent intercellular peptide transfer. **a**, Characterization of A431 cells and A431 cells ectopically expressing connexin 43 (Cx43). Equal numbers of A431 cells or A431 cells stably transfected with an expression construct for human Cx43 (A431/Cx43) were separated by 10% SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and analysed by western blotting. The filter was probed with an anti-Cx43 antibody. The position of the relative molecular mass standards is indicated. **b**, Intercellular peptide diffusion. A431/Cx43 cells were grown to subconfluency and one cell was micro-injected with a mixture of the FL-peptide (sequence DRLDRLDR[C-fluorescein]) and dextran-TxR (M_r 70K). Living cells were analysed 30 min after micro-injection and the distribution of peptide (left panel) and dextran (middle panel) were determined. The right-hand panel shows a transmission image of these cells. FL-peptides move from the micro-injected cell into surrounding cells. **c**, Cx43 expression and intercellular peptide diffusion. Using the protocol depicted in Fig. 1b, peptide diffusion was quantified in A431 cells, A431/Cx43 cells and A431/Cx43 cultured in the presence of the inhibitor 2-APB. Peptide diffusion to neighbouring cells was observed only in the Cx43-containing cells with open gap junctions.

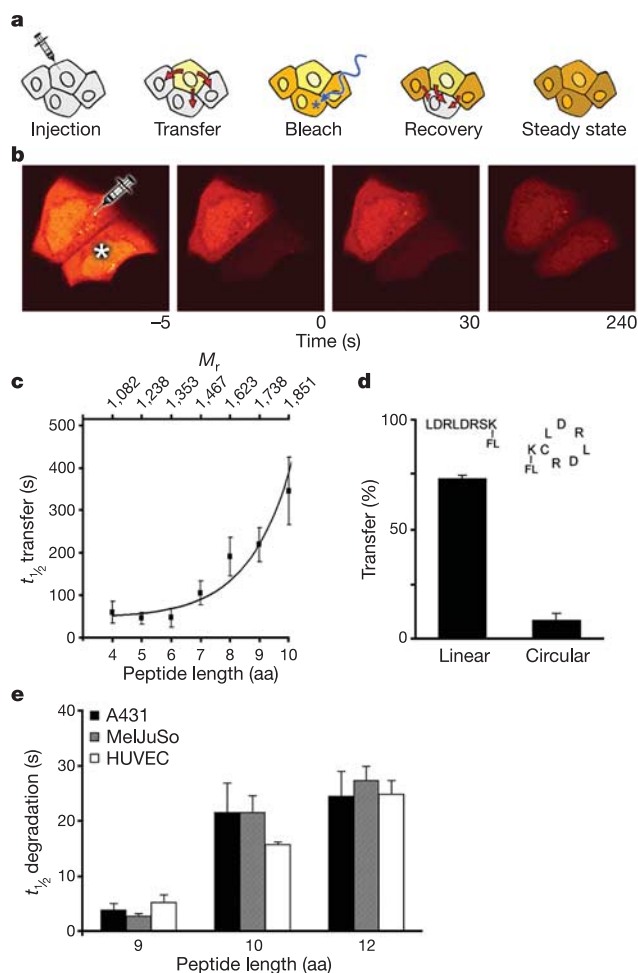


Figure 2 Size and structure dependency for Cx43-mediated intercellular peptide transfer. **a**, Experimental procedure. A431/Cx43 cells were seeded at a high dilution to grow small islands of cells. One cell was micro-injected with a mixture of FL-peptide and dextran-TxR (M_r 70K) as micro-injection marker. After redistribution of the FL-peptide over the cells, fluorescence in one cell was ablated by photobleaching and recovery of fluorescence in the cell was monitored over time to determine the relationship between size, structure and gap-junction-mediated intercellular peptide transfer. **b**, Time-lapse experiment to measure intercellular peptide transfer. An island of two A431/Cx43 cells was assayed. A 10-mer FL-peptide was micro-injected in the top-left cell, which distributed equally over the two cells. The fluorescence in the bottom-right cell (indicated by an asterisk) was photo-inactivated (bleached) and recovered over time to the cost of fluorescence in the micro-injected cell. (See Supplementary Movie.) **c**, Peptide length and kinetics of Cx43-dependent peptide transfer. Following the protocol depicted in **a**, the rate of reappearance of fluorescence in the photobleached cell was quantified for protected fluorescein-labelled peptides of four to ten amino acids (aa). The time of recovery of half the maximal fluorescence ($t_{1/2}$) is depicted ($n > 6$ measurements per data point). Relative molecular mass of the peptides is indicated at the top. **d**, Secondary structure and Cx43-dependent peptide transfer. An 8-mer protected FL-peptide was synthesized as linear or circular peptide. Peptides were introduced to A431/Cx43 by micro-injection and peptide transfer was determined as in Fig. 1c. **e**, The half-life of intracellular peptides. Internally quenched 8-mer peptides with N-terminal extensions of one, two and four amino acids were micro-injected into A431, MeJuSo or HUVEC cells and the degradation was detected through the appearance of fluorescence over time, as described³. Shown is the peptide half-life ($n > 8$ measurements per data point). N-terminal extensions increase peptide half-life and no major differences are observed between the various cell lines. Error bars in **c–e** indicate s.d.

determined the intracellular half-life of peptides by micro-injecting internally quenched fluorescent peptides. These peptides become fluorescent upon degradation, due to spatial separation of quencher and fluorophore¹⁵. Internally quenched peptides of various lengths were micro-injected in A431 cells, a melanoma cell line (MelJuSo) or primary human umbilical cord endothelial cells (HUVEC) and the half-life was determined. In line with previous results¹⁵, a 10- or 12-mer peptide had a longer half-life in the cytoplasm than a 9-mer peptide but no significant differences were observed between the cell lines (Fig. 2e).

By comparing the half-life of peptides *in vivo* and the rate of intercellular peptide transfer, the efficiency of intercellular peptide transfer in A431/Cx43 cells can be estimated. Less than 2% of intracellular peptides are handled by the peptide transporter TAP^{15,17} and gap-junction-mediated intercellular peptide transfer is about equally inefficient because most peptides are destroyed by amino peptidases. In neighbouring cells most of the transferred peptides will also be degraded by cytosolic amino peptidases, allowing transfer of only few peptides into the next cell layer. Cytosolic peptidase activity will therefore limit, but not prevent, spreading of peptides to adjacent cells, thus restricting cross-presentation by innocent bystander cells.

To investigate whether immunologically relevant peptides can be cross-presented, we expressed a green fluorescent protein (GFP)-tagged-ubiquitin (Ub)-influenza matrix peptide (GFP-Ub-influenza matrix(57–65), or FluM_{57–65}) chimera in A431/Cx43 cells. The FluM 9-mer peptide is released from GFP-Ub by ubiquitin hydrolase activity in the cytosol and can be recognized by a specific T-cell clone only when presented by HLA-A2 molecules. The peptide-expressing wild-type A431 and A431/Cx43 cells were co-cultured overnight with HLA-A2 transfectants, and subsequently an HLA-A2-restricted FluM_{57–65}-specific T-cell clone was added (Fig. 3a, b). Because A431 cells do not endogenously express HLA-A2, activation of the FluM-specific T-cell clone can only occur following gap-junctional transfer of the peptide between the peptide-expressing (donor) and the HLA-A2-expressing (acceptor) cells. A T-cell response resulting in interferon- γ (IFN- γ) secretion was observed only when both the peptide-expressing donor and the HLA-A2-expressing acceptor cells express gap junctions (which is required for a functional channel) and no significant response is found for the other conditions (Fig. 3c). The same experimental conditions were used to test whether primary HUVEC endothelial cells could cross-present peptides. HLA-A2-negative HUVEC were transfected with the GFP-Ub-FluM_{57–65} construct and co-cultured with HLA-A2-negative or HLA-A2-positive HUVEC cells. To determine the relative efficiency of direct presentation, HLA-A2-positive HUVEC cells were transfected with the same construct followed by identical co-culture (Fig. 3d). HLA-A2-restricted presentation of FluM_{57–65} peptide generated by the neighbouring cells still strongly stimulated specific T cells when compared with direct presentation. These experiments indicate that antigenic epitopes can be cross-presented after transfer through gap junctions to neighbouring cells, resulting in innocent bystander recognition.

For the initiation of a CTL response, professional APCs have to present antigenic peptides to T cells in the secondary lymphoid organs in a process called cross-priming. For this, professional (mobile) APCs have to acquire antigenic peptides from tissue cells, possibly by gap-junction-mediated peptide transfer. To test this hypothesis we stained human epidermis with antibodies against MHC class II to detect professional APCs (Langerhans cells) and antibodies against Cx43. Characteristic MHC class II-positive Langerhans cells expressed high levels of Cx43 compared with the surrounding keratinocytes (Fig. 4a). Staining of sections from human appendix (intestine) also revealed Cx43 gap junctions between MHC class II-positive (probably dendritic) cells and the surrounding tissue (Fig. 4a). These gap junctions facilitate the

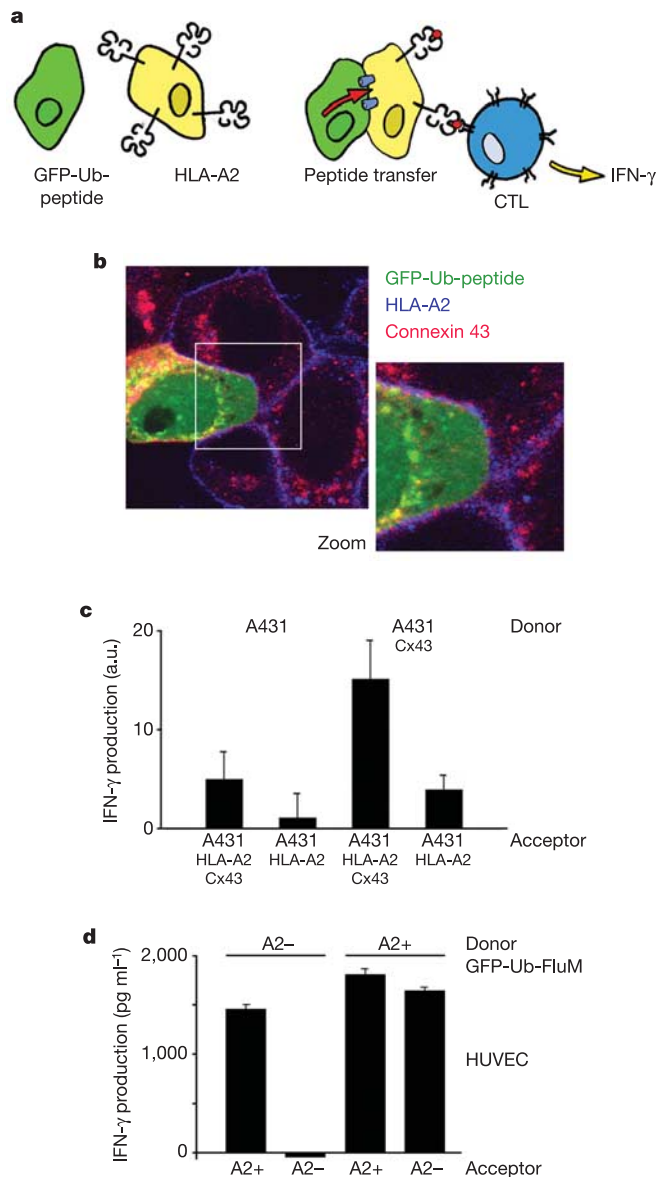


Figure 3 Gap-junction-mediated intercellular peptide transfer for cross-presentation by HLA-A2 molecules. **a**, Experimental procedure. HLA-A2-negative donor cells are transfected with a construct expressing a GFP-ubiquitin-FluM peptide chimera. The donor cells were co-cultured with HLA-A2-positive acceptor cells and peptide transfer and presentation was assayed using HLA-A2-restricted FluM-specific CTL measuring IFN- γ secretion. **b**, Three-colour analysis of the co-culture experiment. A431/Cx43 cells transfected with the GFP-Ub-FluM_{57–65} construct were co-cultured with HLA-A2-transfected A431/Cx43 cells. The co-culture was stained with anti-HLA-A2 antibodies (blue) and anti-Cx43 antibodies (red) and analysed by CLSM. A close-up of the indicated region is shown on the right. Gap junctions are seen at the contact site between the GFP- and the HLA-A2-expressing cells. Intracellular Cx43 staining is the result of internalized Cx43 that failed to assemble into gap junctions. **c**, Gap-junction-dependent peptide cross-presentation by A431/Cx43 cells. GFP-Ub-FluM_{57–65} peptide-expressing A431 or A431/Cx43 donor cells were co-cultured with HLA-A2-expressing A431 or A431/Cx43 acceptor cells and specific CTL activation was detected through IFN- γ production. Specific CTL activation was only detected when both donor and acceptor cells expressed Cx43, which is required for a functional gap junction. **d**, Peptide cross-presentation by primary HUVEC cells. Isolated HLA-A2-positive or -negative HUVECs were transfected with the GFP-Ub-FluM_{57–65} peptide construct. These donor cells were co-cultured with again HLA-A2-positive or -negative HUVEC acceptor cells and activation of specific HLA-A2-restricted CTLs was measured through IFN- γ production. Peptides produced in HLA-A2-negative donor cells were presented by HLA-A2-positive acceptor cells. Error bars in **c** and **d** indicate s.e.m.

transfer of many small molecules, but they might also enable the APCs to sample a fraction of the peptide content of surrounding cells.

To test whether professional APCs are able to contact infected cells and acquire peptides by gap-junctional contact, HLA-A2-positive human primary monocytes were activated by IFN- γ and TNF- α to induce Cx43 expression¹⁸. At the same time, HLA-A2-negative A431, A431/Cx43 or HUVEC cells were infected with influenza. After 16 h of infection, influenza propagation was inhibited and the infected cells were loaded with the dye calcein-AM that diffuses through gap junctions. Subsequently, the activated monocytes were co-cultured with the influenza-infected cells for 18 h. Monocytes making gap-junctional contact with the dye-loaded cell population acquired the calcein dye and were separated from the dye-negative monocytes from the same culture by fluorescence-activated cell sorting (FACS) (Fig. 4b, c). Both monocyte populations were strongly adhering to the monolayer and collection required trypsin treatment. This protocol also showed calcein transfer from HUVEC into monocyte-derived dendritic cells but not between dye-loaded apoptotic bodies from A431/Cx43 and intact A431/Cx43 cells (not shown). Presentation of the FluM₅₇₋₆₅ peptide in the context of HLA-A2 was subsequently assayed for both monocyte populations by CTL (Fig. 4d). Because gap-junction-negative A431 cells only revealed monocytes with background fluorescence, T-cell stimulation was determined only for the dye-negative population. A CTL response was detected only for the dye-

containing monocytes, implying that these monocytes presented an influenza peptide generated in HUVEC and A431/Cx43 cells (Fig. 4d). Cx43 is upregulated in human monocytes upon stimulation with IFN- γ and TNF- α ¹⁸, and non-stimulated monocytes did not cross-present antigens in our protocol (not shown). The monocytes apparently obtained these peptides after gap-junctional contact with the influenza-infected cells, whereas both peptides and dye from A431 cells were not transferred to monocytes because of the absence of gap-junctional contact. To control for any effect of gap-junctional communication on antigen presentation by monocytes, the experiment was repeated without prior influenza infection. Again, the calcein-positive and -negative monocyte populations were isolated and pulsed with FluM₅₇₋₆₅ peptides before addition of specific CTLs (Fig. 4e). Gap-junctional contact with HUVEC does not affect antigen presentation by monocytes. These data suggest that monocytes are able to form gap junctions with other tissues after receiving a 'danger signal' in the form of IFN- γ and TNF- α . The activated monocytes then sample a blueprint of peptides generated in the infected cells for presentation and cross-priming at other sites.

Our data reveal a previously unknown mechanism of antigenic peptide transfer, in which the peptide diffuses, via gap junctions, from the cytoplasm of one cell directly into the cytoplasm of another and thus to the MHC class I antigen presentation pathway of its neighbour, resulting in CTL recognition of these innocent bystanders. Consequently, neighbouring cells can be primed with a

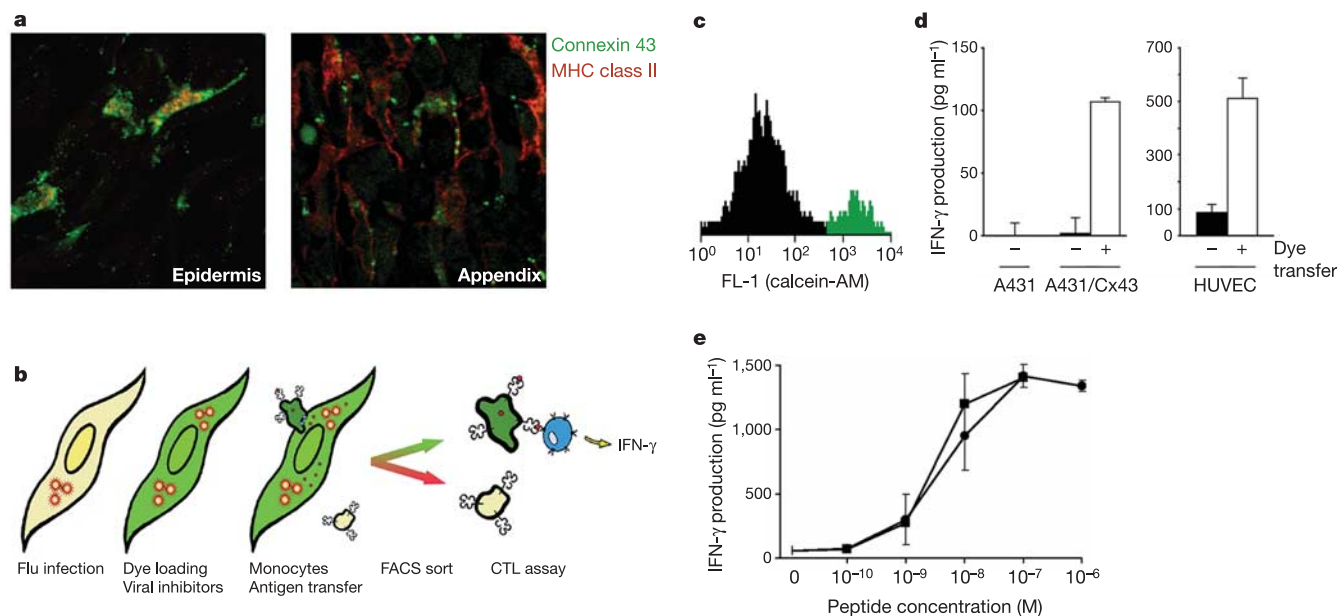


Figure 4 Gap-junction-mediated transfer of influenza virus antigens to APCs and CTL stimulation. **a**, Antigen-presenting cells and gap junctions in human tissue. Human epidermis was sectioned and stained with anti-MHC class II antibodies (red) to detect Langerhans cells, and anti-Cx43 antibodies (green) to detect gap junctions (left). Human appendix was sectioned and stained as described above. MHC class II-positive cells also make Cx43 contacts with the surrounding tissue (right). **b**, Experimental procedure. HLA-A2-negative donor cells were infected with influenza virus and loaded with the green dye calcein-AM before co-culture with HLA-A2-positive monocytes activated for 18 h with IFN- γ and TNF- α (the acceptor cells). Calcein will diffuse through gap junctions to other cells including monocytes. After 18 h of co-culture, calcein-positive and -negative monocytes were isolated from the same culture and both populations were assayed for activation of HLA-A2-restricted FluM₅₇₋₆₅-specific CTLs as detected by IFN- γ secretion. **c**, FACS profiles of the calcein-positive (green) and -negative (black) IFN- γ - and TNF- α -activated monocytes isolated from HUVEC cells after 18 h co-culturing. Most monocytes

fail to acquire calcein from the donor HUVEC monolayer and can be clearly separated from the dye-positive population. **d**, Cross-presentation by human professional APCs. The donor cells A431, A431/Cx43 and HUVEC were infected with influenza virus and loaded with calcein-AM before co-culture with the HLA-A2-positive monocytes, as described in **b**. The acceptor monocytes were first activated with IFN- γ and TNF- α to induce expression of Cx43. Calcein-positive and -negative monocytes were isolated by FACS. An HLA-A2-restricted CTL measured presentation of the FluM₅₇₋₆₅ epitope by monocytes. Because A431 did not reveal dye-positive monocytes, only dye-negative monocytes were assayed. **e**, Gap-junction contact and antigen presentation qualities. HLA-A2-positive monocytes were co-cultured with non-infected dye-loaded HLA-A2-negative HUVEC cells preloaded with calcein-AM. Calcein-positive (squares) and -negative (circles) monocytes were isolated and loaded with various concentrations of FluM₅₇₋₆₅ peptide before stimulation of the HLA-A2-restricted FluM₅₇₋₆₅-specific CTL was assayed. Error bars in **d** and **e** indicate s.d.

viral peptide, recognized and possibly killed by CTLs before actual infection by the virus or before the production of viral DriPs¹, albeit at the cost of some cells that would not have been infected. However, this will efficiently prevent spread of infection whereby neighbouring cells form a 'cordon sanitaire' surrounding the infected cell. Also, auto-antigenic peptides may be transferred through gap junctions, which could explain the observed recognition of endothelial cells surrounding Islet cells by activated CD8⁺ insulin-specific T cells¹⁹.

In principle, tissue interconnected by gap junctions may be eliminated after antigenic peptide transfer. However, only a few peptides will successfully enter neighbouring cells and peptidase activity will destroy the 'visiting' peptides, restricting further diffusion into another cell layer. Consequently, antigens have to be expressed at relatively high levels to ensure antigen transfer to APCs for cross-presentation to occur, as has been previously suggested²⁰.

Cross-presentation of antigenic peptides by APCs is required for expansion of a specific T-cell population. So far only a few mechanisms explaining cross-presentation have been described²⁻⁴. Transfer of antigenic proteins from a cell into an APC may occur following protein secretion and uptake by dendritic cells. Exogenous (stable) proteins and heat-shock protein-associated antigens²¹ can induce CD8⁺ T-cell responses but these proteins first have to enter either the general MHC class I presentation pathway, or be presented on recycling MHC class I molecules²². Antigens coupled to beads can be transferred to the cytoplasm in early phagosomes of dendritic cells²³⁻²⁵. Apoptotic bodies are a possible physiological equivalent for beads and are cross-presented by CD8⁺ DCs in the mouse (this DC subset has not been identified in man)⁴. Cross-presentation then would require proteins to be released from the apoptotic bodies during an early phase of uptake²³. Because many tumours as well as viruses express anti-apoptotic proteins, their antigens will not induce a proper T-cell response. The presence or absence of gap junctions in antigen-expressing cells may explain why some have reported that antigenic proteasomal degradation products can be transferred from antigen donor cells to acceptor cells²⁶, and others that such fragments are excluded from cross-presentation²⁷⁻²⁹.

We present here an alternative mechanism for cross-presentation: gap-junction-mediated immunological coupling (GMIC). This mechanism is fundamentally different from those proposed until now and allows direct antigenic fragment transfer between the cytoplasm of cells. Peptides can probably be transferred through many different gap junctions but haematopoietic cells only express Cx43. Monocytes upregulate Cx43 after receiving 'danger signals' in the form of IFN- γ and TNF- α or lipopolysaccharide¹⁸. By establishing gap-junctional contact with local cells, they may sample a fraction of peptides expressed in these cells and transfer their antigenic peptides to lymph nodes for T-cell activation and expansion. Tissue-specific APCs also seem to form Cx43 gap junctions with surrounding cells. Again, through immunological coupling these tissue-specific APCs may sample peptides from their environment for cross-presentation to T cells at other locations.

Gap-junction-mediated immunological coupling can result in innocent bystander recognition controlled by cytosolic amino peptidase activity. Immunological coupling allows a very quick CTL response against cells at high risk of infection. Cross-presentation is analogous to presentation by bystander cells when the acceptor cell is a professional APC. This may be a crucial method of ensuring proper T-cell responses against viruses hiding in cells and cells protected from undergoing apoptosis through anti-apoptotic viral proteins. Antigen acquisition by gap junctions couples the antigen presentation pathways of neighbouring cells, resulting in cross-presentation by innocent bystanders and APCs. □

Methods

Cells, antibodies and peptides

The following antibodies were used: mAb MA2.1 and BB7.2 (anti-HLA-A2), mAb 1B5 (anti-HLA-DR); rabbit anti-Cx43 (Sigma). Fluorescent secondary antibodies were from Molecular Probes.

A431 cells were stably transfected with human Cx43 complementary DNA in pcDNA3 (provided by B. Giepmans). A431 and A431/Cx43 cells were stably transfected with HLA-A2.01 cDNA cloned in pcDNA3. Human monocytes were obtained from healthy HLA-typed volunteers by isolation of peripheral blood lymphocytes (PBL) by Ficoll centrifugation and a subsequent short adherence step. Monocytes were activated by an 18-h culture in the presence of IFN- γ (1 ng ml⁻¹) and TNF- α (1 ng ml⁻¹), as described¹⁸. HUVEC cells were isolated following standard protocols. Both monocytes and HUVECs were tested for HLA-A2 expression with the mAb BB7.2. The human T-cell clone (InfA13TGA) overexpressing telomerase is recognizing FluM₅₇₋₆₅ in the context of HLA-A2 (ref. 30).

A431/Cx43 cells expressing GFP-Ub-FluM₅₇₋₆₅ were co-cultured with A431/Cx43 cells expressing HLA-A2, and stained with MA2.1, followed by formaldehyde fixation and staining with anti-Cx43 antiserum and secondary TexasRed- and Cy5-labelled antibodies. Paraffin-embedded sections of human epidermis and appendix were stained with the mAb 1B5 and Cx43 antiserum followed by TexasRed-, Cy5- or fluorescein-labelled secondary antibodies. No fluorescence was detected in control stainings (secondary antibody only). Images were made with a Leica TCS-SP2 CLSM. Peptides were synthesized by Fmoc (fluorenylmethoxycarbonyl) chemistry, purified and confirmed by mass spectrometry. FluM₅₇₋₆₅ peptide sequence: GILGFVFTL. Peptides for intercellular peptide transfer experiments were composed of D-amino acids. An N-terminal naftylsulphonyl group and C-terminal amidation were introduced to prevent degradation¹⁵. Fluorescein was conjugated to cysteine residues using fluorescein-5-iodoacetamide (Molecular Probes). The sequences were: LDRLDRLDR[C-fluorescein] and N-terminal truncations down to LDR[C-fluorescein]. The circular peptide was prepared by extending the sequence LDRLDRCCK with a bromo-acetyl moiety. Cyclization (thioether formation by reaction of cysteine with the bromo-acetyl moiety) was performed at pH 8, followed by fluorescein isothiocyanate (FITC) conjugation of lysine. Internally quenched peptides were synthesized as described: with sequences T[K-Dabcy1]NKTER[C-fluorescein]Y with either an N-terminal P or LGP addition for the 10- and 12-mer peptide.

Peptide transfer and degradation

To detect intercellular peptide transfer, cells were seeded at a high dilution on coverslips resulting in small cell islands after 2-3 days of culture. One cell was micro-injected with a mixture of FL-peptide and TxR-dextran (M_r 70K; Molecular Probes) in the presence or absence of 100 μ M 2-aminoethoxydiphenylborate (2-APB, Sigma). Quantification was performed 30-90 min after micro-injection. To determine the rate of intercellular peptide transfer, the identical experiment was performed followed by subsequent bleaching of one of the FL-peptide acceptor cells with 488-nm laser light. FRAP experiments were performed with a Leica TCS-SP2 CLSM, as described¹⁵. Peptide degradation of internally quenched peptides was assayed as described¹⁵.

Cross-presentation assays

The GFP-Ub-FluM₅₇₋₆₅ construct was made by PCR on ubiquitin using a reverse primer with a flanking region containing the sequence encoding the FluM₅₇₋₆₅ peptide, placing the peptide directly adjacent into the ubiquitin-splicing site following Gly₇₆ of ubiquitin. This product was cloned in pEGFP-C1 (Clontech).

A431 or A431/Cx43 cells were transfected by electroporation using a Biorad Gene Pulser. HUVEC cells were transfected using the Amara HUVEC Nucleofector kit. 24 h after transfection, cells were FACS sorted for equal GFP expression and the GFP-positive cells were co-cultured with equal numbers of acceptor cells for another 24 h followed by the addition of T-cell clone InfA13TGA. After 18 h co-culture, IFN- γ secretion was measured by enzyme-linked immunosorbent assay (ELISA) (PeliKine Compact, CLB) and quantified using internal standards.

To study antigen transfer, A431, A431/Cx43 or (HLA-A2-negative) HUVEC cells were infected with influenza A virus strain A/WSN/33 (H1N1) at a concentration of 4×10^6 plaque-forming units (PFU) per ml for 18 h. After 16 h, propagation of infection was inhibited by 10 μ g ml⁻¹ ribavirin (ICN) and 25 μ g ml⁻¹ chloroquine. These inhibitors were present during all subsequent culture steps. Cells were subsequently labelled with calcein-AM (Molecular Probes) at a concentration of 1 μ g ml⁻¹ for 45 min followed by extensive washing. No post-wash labelling of monocytes was detected under these conditions. Activated HLA-A2-positive monocytes were added to the culture (in approximately a twofold cell excess). Cells were co-cultured for 18 h and removed by trypsinization because both populations were strongly adhering to the underlying monolayer. FACS was subsequently used to isolate the calcein-positive and -negative monocytes. About 10-20% calcein-positive monocytes were recovered from HUVEC co-cultures and 5-8% from A431/Cx43. A431 cells revealed monocytes with only background fluorescence. Calcein-positive and -negative monocytes were subsequently co-cultured for another 18 h with the CTL clone and IFN- γ secretion was determined by ELISA assay.

To control for antigen presentation capacity, HLA-A2-expressing monocytes were co-cultured with HUVEC cells without influenza infection and inhibitors. Calcein-positive and -negative monocytes were isolated by FACS and pulsed with different concentrations of the FluM₅₇₋₆₅ peptide before co-culture with the CTL clone for another 18 h and detection of IFN- γ secretion by ELISA.

Received 25 August; accepted 14 December 2004; doi:10.1038/nature03290.

1. Yewdell, J. W., Reits, E. & Neefjes, J. Making sense of mass destruction: quantitating MHC class I antigen presentation. *Nature Rev. Immunol.* **3**, 952-961 (2003).

2. Rock, K. L. The ins and outs of cross-presentation. *Nature Immunol.* **4**, 941–943 (2003).
3. Ackerman, A. L. & Cresswell, P. Cellular mechanisms governing cross-presentation of exogenous antigens. *Nature Immunol.* **5**, 678–684 (2004).
4. Heath, W. R. *et al.* Cross-presentation, dendritic cell subsets, and the generation of immunity to cellular antigens. *Immunol. Rev.* **199**, 9–26 (2004).
5. Segretain, D. & Falk, M. M. Regulation of connexin biosynthesis, assembly, gap junction formation, and removal. *Biochim. Biophys. Acta* **1662**, 3–21 (2004).
6. Goldberg, G. S., Valiunas, V. & Brink, P. R. Selective permeability of gap junction channels. *Biochim. Biophys. Acta* **1662**, 96–101 (2004).
7. Warn-Cramer, B. J. & Lau, A. F. Regulation of gap junctions by tyrosine protein kinases. *Biochim. Biophys. Acta* **1662**, 81–95 (2004).
8. Oviedo-Orta, E. & Evans, W. H. Gap junctions and connexin-mediated communication in the immune system. *Biochim. Biophys. Acta* **1662**, 102–112 (2004).
9. Goldberg, G. S. & Lau, A. F. Dynamics of connexin43 phosphorylation in pp60v-src-transformed cells. *Biochem. J.* **295**, 735–742 (1993).
10. Esinduy, C. B., Chang, C. C., Trosko, J. E. & Ruch, R. J. *In vitro* growth inhibition of neoplastically transformed cells by non-transformed cells: requirement for gap junctional intercellular communication. *Carcinogenesis* **16**, 915–921 (1995).
11. Fischer, N. O., Mbuy, G. N. & Woodruff, R. I. HSV-2 disrupts gap junctional intercellular communication by mammalian cells *in vitro*. *J. Virol. Methods* **91**, 157–166 (2001).
12. Oelze, I., Kartenbeck, J., Crusius, K. & Alonso, A. Human papillomavirus type 16 E5 protein affects cell-cell communication in an epithelial cell line. *J. Virol.* **69**, 4489–4494 (1995).
13. Mesnil, M. & Yamasaki, H. Bystander effect in herpes simplex virus–thymidine kinase/ganciclovir cancer gene therapy: role of gap-junctional intercellular communication. *Cancer Res.* **60**, 3989–3999 (2000).
14. Giepmans, B. N., Hengeveld, T., Postma, F. R. & Moolenaar, W. H. Interaction of c-Src with gap junction protein connexin-43. Role in the regulation of cell-cell communication. *J. Biol. Chem.* **276**, 8544–8549 (2001).
15. Reits, E. *et al.* Peptide diffusion, protection, and degradation in nuclear and cytoplasmic compartments before antigen presentation by MHC class I. *Immunity* **18**, 97–108 (2003).
16. Harks, E. G. *et al.* Besides affecting the intracellular calcium signaling, 2-APB reversibly blocks gap junctional coupling in confluent monolayers. *FASEB J.* **17**, 941–943 (2003).
17. Princiotto, M. F. *et al.* Quantitating protein synthesis, degradation, and endogenous antigen processing. *Immunity* **18**, 343–354 (2003).
18. Eugenin, E. A., Branes, M. C., Berman, J. W. & Saez, J. C. TNF- α plus IFN- γ induce connexin43 expression and formation of gap junctions between human monocytes/macrophages that enhance physiological responses. *J. Immunol.* **170**, 1320–1328 (2003).
19. Savinov, A. Y., Wong, F. S., Stonebraker, A. C. & Chervonsky, A. V. Presentation of antigen by endothelial cells and chemotaxis are required for homing of insulin-specific CD8⁺ T cells. *J. Exp. Med.* **197**, 643–656 (2003).
20. Kurts, C., Heath, W. R., Carbone, F. R., Kosaka, H. & Miller, J. F. Cross-presentation of self antigens to CD8⁺ T cells: the balance between tolerance and autoimmunity. *Novartis Found. Symp.* **215**, 172–181 (1998).
21. Li, Z., Menoret, A. & Srivastava, P. Roles of heat-shock proteins in antigen presentation and cross-presentation. *Curr. Opin. Immunol.* **14**, 45–51 (2002).
22. Gromme, M. *et al.* Recycling MHC class I molecules and endosomal peptide loading. *Proc. Natl Acad. Sci. USA* **96**, 10326–10331 (1999).
23. Gil-Torregrosa, B. C. *et al.* Control of cross-presentation during dendritic cell maturation. *Eur. J. Immunol.* **34**, 398–407 (2004).
24. Guernonprez, P. *et al.* ER-phagosome fusion defines an MHC class I cross-presentation compartment in dendritic cells. *Nature* **425**, 397–402 (2003).
25. Houde, M. *et al.* Phagosomes are competent organelles for antigen cross-presentation. *Nature* **425**, 402–406 (2003).
26. Serna, A., Ramirez, M. C., Soukhanova, A. & Sigal, L. J. Cutting edge: efficient MHC class I cross-presentation during early vaccinia infection requires the transfer of proteasomal intermediates between antigen donor and presenting cells. *J. Immunol.* **171**, 5668–5672 (2003).
27. Wolkers, M. C., Brouwenstijn, N., Bakker, A. H., Toebes, M. & Schumacher, T. N. Antigen bias in T cell cross-priming. *Science* **304**, 1314–1317 (2004).
28. Norbury, C. C. *et al.* CD8⁺ T cell cross-priming via transfer of proteasome substrates. *Science* **304**, 1318–1321 (2004).
29. Shen, L. & Rock, K. L. Cellular protein is the source of cross-priming antigen *in vivo*. *Proc. Natl Acad. Sci. USA* **101**, 3035–3040 (2004).
30. Verra, N. C. *et al.* Human telomerase reverse transcriptase-transduced human cytotoxic T cells suppress the growth of human melanoma in immunodeficient mice. *Cancer Res.* **64**, 2153–2161 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. van Beem and E. Sellink for human monocyte and HUVEC isolations, W. Moolenaar, B. Giepmans and L. van Zeijl for A431/Cx43 cells and Cx43 reagents, R. Luiten and H. Spits for T-cell clone (InfA13TGA), W. E. Benckhuijsen for peptide synthesis, E. Mesman and M. Tjin-A-Koeng for immunohistochemistry, K. Jalink for experimental support, and H. Pickersgill, A. Griekspoor and M. Wolkers for critical reading. This work was supported by grants from the Dutch Cancer Society KWF.

Authors' contributions J.N. and C.H. performed most experiments with support from E.R. Constructs were made by L.J. and peptides by J.W.D. Supervision by J.N.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.N. (J.Neeftjes@nki.nl).

CD4⁺ T-cell help controls CD8⁺ T-cell memory via TRAIL-mediated activation-induced cell death

Edith M. Janssen^{1*}, Nathalie M. Droin^{1*}, Edward E. Lemmens¹, Michael J. Pinkoski¹, Steven J. Bensinger¹, Benjamin D. Ehst², Thomas S. Griffith³, Douglas R. Green^{1*} & Stephen P. Schoenberger^{1*}

¹Division of Cellular Immunology, La Jolla Institute for Allergy and Immunology, 10355 Science Center Drive, San Diego, California 92121, USA

²Department of Microbiology, Center for Immunology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA

³Department of Urology, The Interdisciplinary Graduate Program in Immunology, University of Iowa, Iowa City, Iowa 52242, USA

* These authors contributed equally to this work

The 'help' provided by CD4⁺ T lymphocytes during the priming of CD8⁺ T lymphocytes confers a key feature of immune memory: the capacity for autonomous secondary expansion following re-encounter with antigen^{1–4}. Once primed in the presence of CD4⁺ T cells, 'helped' CD8⁺ T cells acquire the ability to undergo a second round of clonal expansion upon restimulation in the absence of T-cell help. 'Helpless' CD8⁺ T cells that are primed in the absence of CD4⁺ T cells, in contrast, can mediate effector functions such as cytotoxicity and cytokine secretion upon restimulation, but do not undergo a second round of clonal expansion. These disparate responses have features of being 'programmed', that is, guided by signals that are transmitted to naive CD8⁺ T cells during priming, which encode specific fates for their clonal progeny. Here we explore the instructional programme that governs the secondary response of CD8⁺ T cells and find that helpless cells undergo death by activation-induced cell death upon secondary stimulation. This death is mediated by tumour-necrosis factor (TNF)-related apoptosis-inducing ligand (TRAIL). Regulation of *Trail* expression can therefore account for the role of CD4⁺ T cells in the generation of CD8⁺ T cell memory and represents a novel mechanism for controlling adaptive immune responses.

As a first step towards identifying the defect present in helpless CD8⁺ T cells, we examined their capacity to initiate a secondary proliferative response upon restimulation. This was done using an *in vivo* cross-priming system that features immunization with a cellular vaccine expressing an H-2D^b-restricted model antigen, E1B_{192–200} (TAP^{−/−} Ad5E1-MEC)⁵. Seven days after *in vivo* priming, CD8⁺ T cells were obtained and labelled with the viable fluorescent dye CFSE (5, 6-carboxy-succinimidyl-fluorescein-ester) to allow visualization of the clonal expansion⁶. CFSE dilution was assessed in E1B_{192–200}-specific effector CD8⁺ T cells 4 days after *in vitro* restimulation on TAP^{+/+}Ad5E1-MEC (ref. 5). Effector CD8⁺ T cells from control mice (helped CD8⁺ T cells) both proliferated and accumulated upon secondary encounter with antigen, with CFSE dilution profiles indicating that they underwent multiple rounds of division (Fig. 1a, top left panel). The majority of CD8⁺ T cells obtained from CD4-depleted mice (helpless CD8⁺ T cells), in contrast, either failed to divide or went through a limited number (one or two rounds) of divisions (Fig. 1a, top right panel). The absolute number of E1B_{192–200}-specific cells in these cultures decreased, suggesting that these helpless cells were dying following restimulation. Addition of caspase inhibitors quinoline-Val-Asp-CH₂-difluorophenoxy (qVD-OPh) and benzyloxycarbonyl-Val-Ala-Asp-fluoromethyl ketone (zVAD-fmk) restored secondary expansion in the purified helpless CD8⁺ T cells, and resulted in comparable proliferation and accumulation to that observed in the helped CD8⁺ T cells (Fig. 1a, b). These data show that helpless

CD8⁺ T cells have the potential to proliferate and survive upon secondary stimulation, but this outcome may be blocked or suppressed by the induction of caspase-mediated activation-induced cell death (AICD).

To gain insight into the molecular pathways underlying the disparate secondary responses of helped and helpless CD8⁺ T cells, a set of apoptosis-related candidate genes was analysed in the restimulated CD8⁺ T cells by real-time polymerase chain reaction with reverse transcription (RT-PCR). Notably, messenger RNAs for *Bcl-2*, *Bcl-xL* (also known as *Bcl2l1*) and *FLIP*, encoding anti-apoptotic proteins, were selectively upregulated upon activation of helped CD8⁺ T cells (Fig. 2a), cells that subsequently expanded and accumulated (data not shown). In contrast, mRNAs for *FasL* (also known as *Tnfsf6*) and *Trail* (also known as *TNFS10* and *Apo-2L*)^{7,8}, both of which encode pro-apoptotic proteins, were selectively upregulated in helpless CD8⁺ T cells (Fig. 2a), cells which subsequently failed to expand (data not shown). Both purified helped and helpless CD8⁺ T cells upregulated mRNA for *TNF* and TRAIL receptor *DR5* (TRAIL-R2)^{9–11} following restimulation (Fig. 2a).

To further delineate the contribution of these genes, we analysed primary and secondary CD8⁺ T-cell responses in a set of genetically manipulated mice in which *Bcl-2* and *Bcl-xL* are constitutively expressed, or in which TNF receptor (TNFR), Fas and TRAIL expression is absent or functionally deficient. The frequency of E1B_{192–200}-specific CD8⁺ T cells in control and CD4-depleted mice was determined directly *ex vivo* (see Supplementary Fig. S1 for enumeration of primary IFN- γ ⁺ CD8⁺ effectors) and their capacity for secondary expansion was assessed following restimulation *in vitro*. In each of the mouse strains tested, helped CD8⁺ T cells underwent a similar degree of secondary expansion upon restimulation (Fig. 2b). However, neither enforced expression of *Bcl-2* or *Bcl-xL* nor the absence of Fas (*lpr/lpr*) or TNFR (*TNFR1*^{−/−}) rescued secondary expansion in helpless CD8⁺ T cells (Fig. 2b). Effector T cells from *Trail*-deficient mice, in contrast, did not

exhibit the helpless phenotype when primed in the absence of CD4⁺ T cells, and instead underwent substantial secondary expansion following restimulation (Fig. 2b). This result was also observed using a second cross-priming system in which CD8⁺ T cells specific for a different antigen were enumerated physically using peptide/MHC tetramers (Supplementary Fig. S2). In this system, mice were immunized with splenocytes from act-mOVA/K^b−/− mice that express a membrane-bound form of the model antigen chicken ovalbumin (OVA), but lack expression of the H-2K^b MHC class I molecule required for presentation of the immunodominant OVA epitope (OVA_{257–264}) to CD8⁺ T cells¹². Taken together, these results suggest that secondary expansion of helpless CD8⁺ T cells is constrained by TRAIL.

To determine whether TRAIL exerted a similar effect *in vivo*, intact and CD4-depleted wild-type and *Trail*^{−/−} mice were immunized with act-mOVA/K^b−/− splenocytes, with some groups challenged again 10 days later. In this prime/boost setting, the helpless OVA_{257–264}-specific CD8⁺ T cells in *Trail*^{−/−} mice underwent secondary expansion whereas those from wild-type mice did not (Fig. 2c). These results reveal that regardless of whether restimulation occurs *in vitro* or *in vivo*, helpless CD8⁺ T cells lacking *Trail* expression undergo secondary expansion. We further investigated the role of TRAIL using a soluble form of the TRAIL receptor DR5 (DR5-Fc) to block the action of TRAIL during secondary expansion. Addition of DR5-Fc restored secondary expansion of helpless CD8⁺ T cells from wild-type mice (Fig. 2d). Addition of the control fusion proteins Fas-Fc or TNFR-Fc, in contrast, had no effect and none of the soluble receptors affected the secondary expansion of purified helped CD8⁺ T cells (Fig. 2d). Taken together, these results indicate that autocrine TRAIL expression prevents the secondary expansion of helpless CD8⁺ T cells.

Helpless, but not helped, CD8⁺ T cells express *Trail* mRNA upon restimulation (Fig. 2a), raising the possibility that *Trail* is differentially regulated at the transcriptional level in these populations. To test this, a 0.4 kb 5′-flanking sequence of the *Trail* gene was used to drive expression of a reporter in primary cells¹³. Following restimulation, real-time RT-PCR was used to detect reporter expression in the rare antigen-specific transfected CD8⁺ T cells (Fig. 2e). Secondary stimulation led to a strong induction of the *Trail* promoter reporter in helpless CD8⁺ T cells whereas no induction was observed in the helped CD8⁺ T cells (Fig. 2e), indicating that the observed increase in *Trail* mRNA in helpless CD8⁺ T cells is transcriptionally regulated. Taken together, these results demonstrate that although CD8⁺ T cells that are primed in the presence or absence of T help undergo primary expansion with comparable kinetics⁴, their progeny undergo fundamentally different programmes of gene expression upon secondary stimulation. Of these genes, *Trail* expression is the main determinant of the outcome of secondary encounter with antigen.

We extended our observations to a well characterized model of viral infection by examining the role of TRAIL in the CD8⁺ T-cell response to infection with lymphochoriomeningitis virus (LCMV). As in the cross-priming model, CD8⁺ T cells primed in the absence of T help expand and acquire effector function, but are defective in their secondary expansion⁴. Following restimulation, LCMV-specific helped CD8⁺ T cells upregulated mRNA for anti-apoptotic genes including *Bcl-2*, *Bcl-xL* and *FLIP*, but not *Trail* mRNA, whereas the helpless CD8⁺ T cells failed to upregulate these anti-apoptotic genes and induced the expression of *Trail* mRNA (Fig. 3a). Selective induction of *Trail* expression was also observed when helped and helpless GP_{33–41}/K^b-specific CD8⁺ T cells were isolated using GP_{33–41}/K^b tetramers (Supplementary Fig. S3). Expression of *DR5* mRNA, however, was induced in both helped and helpless CD8⁺ T cell populations (Fig. 3a). Secondary expansion of the helpless CD8⁺ T cells was restored when DR5-Fc or zVAD-fmk were added to the cultures, but not in the presence of Fas-Fc (Fig. 3b), and the *Trail* promoter reporter was selectively induced in the

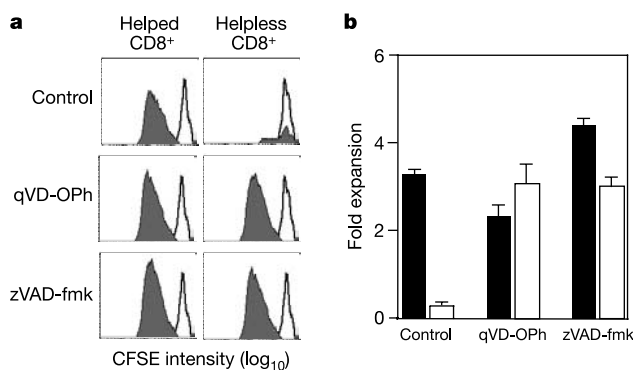


Figure 1 Caspase inhibitors restore secondary expansion in helpless CD8⁺ T cells. **a**, CFSE dilution profile of IFN- γ -producing E1B_{192–200}-specific CD8⁺ T cells from intact (helped) and CD4-depleted (helpless) mice immunized 7 days earlier with 10⁷ irradiated Tap^{−/−}Ad5E1-MEC. CFSE levels in E1B_{192–200}-specific IFN- γ -producing effectors were measured directly *ex vivo* (open histogram) and after a 4-day stimulation *in vitro* (shaded histogram) in the presence of caspase inhibitors qVD-OPh and zVAD-fmk. Results show representative histograms from two independent experiments using 4–5 mice per group. **b**, The secondary expansion of IFN- γ -producing E1B_{192–200}-specific CD8⁺ T cells from intact (helped) and CD4-depleted (helpless) mice immunized 7 days earlier with 10⁷ irradiated Tap^{−/−}Ad5E1-MEC was determined by intracellular IFN- γ staining 6 days after *in vitro* restimulation with Tap^{+/+}Ad5E1-MEC in the presence or absence of the indicated caspase inhibitors. Expansion of E1B_{192–200}-specific helped (filled bars) and helpless (open bars) CD8⁺ T cells was calculated as the fold increase in the absolute number of specific CD8⁺ T cells. Data are shown as mean $\pm$ s.e.m. (n = 4–5 mice per group).

restimulated helpless CD8⁺ T cells (Fig. 3c). These data reveal that in both the cross-priming and viral infection models, *Trail* regulates secondary expansion of CD8⁺ T cells.

Sensitivity to TRAIL-mediated apoptosis is regulated at different levels, including expression of receptor and signalling components, antagonists of receptor signalling, and differences in expression or activation of caspases and their inhibitors¹⁴. In both the cross-priming and LCMV models, helpless and helped CD8⁺ T cells expressed *DR5* following secondary antigenic stimulation (Figs 2 and 3), but only the helped CD8⁺ T cells upregulated *FLIP* mRNA, encoding a protein that can potentially inhibit TRAIL-mediated apoptosis¹⁵. To determine whether resistance to TRAIL receptor signalling contributes to secondary expansion of helped CD8⁺ T cells, we examined effects of exogenous TRAIL *in vitro*. Addition of recombinant TRAIL to helped CD8⁺ T cell cultures completely inhibited secondary expansion, demonstrating that helped CD8⁺ T cells were sensitive to TRAIL-mediated effects (Fig. 4a). Helpless CD8⁺ T cells from *Trail*-deficient mice, which underwent secondary expansion (Fig. 4b), failed to do so in the presence of recombinant

TRAIL. These results suggest that helped and helpless CD8⁺ T cells are equally sensitive to the effects of TRAIL, and that the differences in their secondary response to antigen are therefore likely to be a consequence of *Trail* expression.

Consistent with this idea, we found that the lack of response of helpless CD8⁺ T cells to secondary stimulation dominates over that of helped CD8⁺ T cells. Helped and helpless CD8⁺ T cells were cultured in a transwell system and the secondary proliferative response of each type in the lower wells was determined (Fig. 4c). Whereas addition of helped CD8⁺ T cells to the upper wells had no effect on the responses of either helped or helpless CD8⁺ T cells in the lower wells, the helpless CD8⁺ T cells effectively inhibited the programmed secondary expansion of the helped CD8⁺ T cells plated in the lower wells. This inhibitory effect was blocked by addition of either DR5-Fc or a neutralizing anti-TRAIL monoclonal antibody to the cultures (Fig. 4d). These experiments indicate that the TRAIL produced by helpless CD8⁺ T cells is soluble, and can therefore potentially act on cells that are not in direct contact with the helpless CD8⁺ T cells.

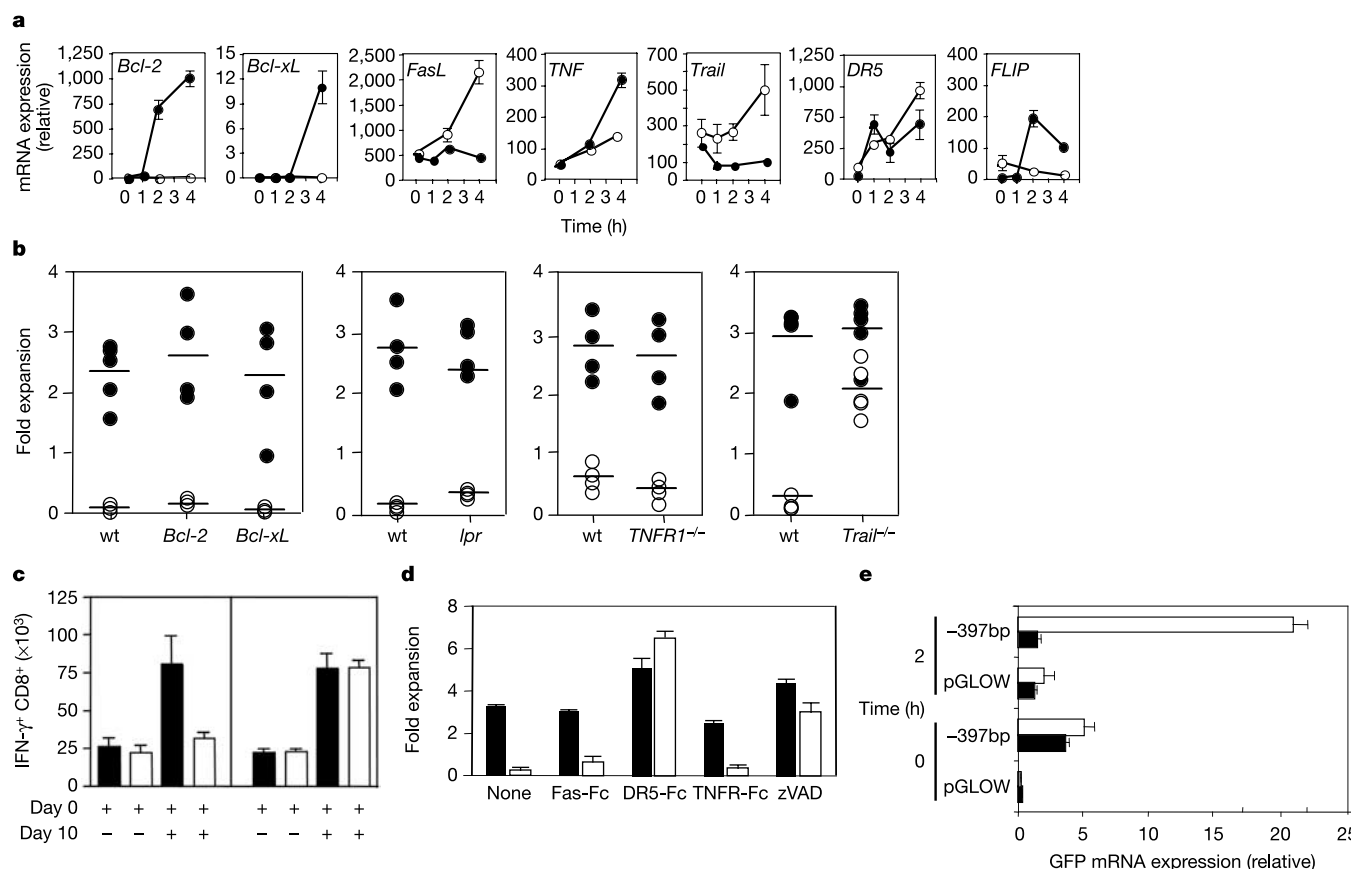


Figure 2 *Trail* expression prevents secondary expansion of helpless CD8⁺ T cells.

a, mRNA levels for the apoptosis-related genes *Bcl-2*, *Bcl-xL*, *FasL*, *TNF*, *Trail*, *DR5* and *FLIP* were determined by real-time RT-PCR at different time points following *in vitro* E1B_{192–200}-peptide restimulation of purified CD8⁺ T cells obtained from intact (filled circle) and CD4-depleted mice (open circle) immunized 7 days earlier with 10⁷ irradiated Tap^{-/-} Ad5E1-MEC. **b**, Intact (filled circle) and CD4-depleted mice (open circle) from indicated strains were immunized with Tap^{-/-} Ad5E1-MEC. The frequency of IFN-γ-producing E1B-specific CD8⁺ T cells was determined directly *ex vivo* and again following a 6-day co-culture with Tap^{+/+} Ad5E1-MEC. The fold expansion of IFN-γ-producing E1B_{192–200}-specific CD8⁺ T cells was calculated as the increase in the absolute number of specific CD8⁺ T cells. **c**, Intact (filled bars) and CD4-depleted mice (open bars) were immunized with 2 × 10⁷ irradiated act-mOVA/K^b-/- splenocytes (day 0) and rechallenged with act-mOVA/K^b-/- splenocytes on day 10. The absolute number of

OVA_{257–264}-specific CD8⁺ T cells per spleen was determined 4 days later by intracellular IFN-γ staining. **d**, Purified CD8⁺ T cells from Tap^{-/-} Ad5E1-MEC-immunized intact (filled bars) and CD4-depleted mice (open bars) were cultured with Tap^{+/+} Ad5E1-MEC in the presence or absence of the indicated soluble death receptors. After 6 days the fold expansion of IFN-γ-producing E1B_{192–200}-specific CD8⁺ T cells was calculated as the increase in the absolute number of specific CD8⁺ T cells. **e**, Purified CD8⁺ T cells from Tap^{-/-} Ad5E1-MEC-immunized intact (filled bars) and CD4-depleted mice (open bars) were cotransfected with empty pGLOW or pGLOW containing the bp -397 of human *Trail* promoter in the presence of the β-galactosidase-expressing vector. Transfected cells were incubated 13 h before stimulation with E1B_{192–200} peptide. GFP mRNA was analysed by real-time RT-PCR and normalized to *18S* and to *LacZ* mRNA (as an efficiency control) expression. Data are shown as mean ± s.e.m. (*n* = 4–5 mice per group).

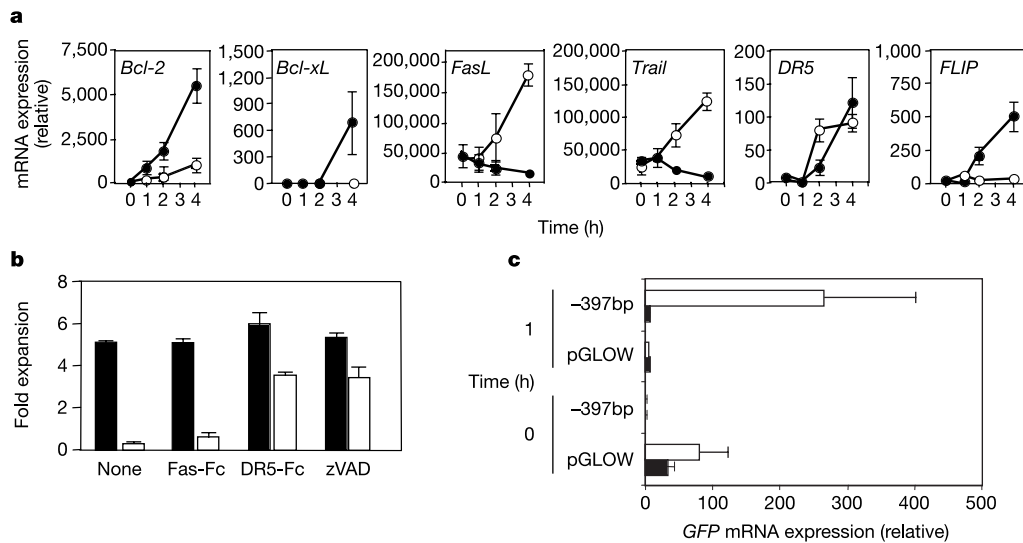


Figure 3 *Trail*-mediated defective secondary expansion of LCMV-specific helpless CD8⁺ T cells. **a**, mRNA levels for *Bcl-2*, *Bcl-xL*, *FasL*, *Trail*, *DR5* and *FLIP* were determined by real-time RT-PCR at different time points following *in vitro* GP₃₃₋₄₁-peptide restimulation of purified CD8⁺ T cells obtained from intact (filled circle) and CD4-depleted mice (open circle) immunized 28 days earlier with 2×10^5 PFU LCMV (Armstrong). **b**, Purified CD8⁺ T cells from immunized intact (filled bars) and CD4-depleted mice (open bars) were cultured with LCMV-infected thioglycollate-induced macrophages in the presence or absence of soluble death receptors. After 6 days the fold expansion was

calculated as the increase in the absolute number of GP₃₃₋₄₁-specific CD8⁺ T cells. **c**, Purified CD8⁺ T cells from intact (filled bars) and CD4-depleted mice infected 28 days earlier (open bars) were cotransfected with empty pGLOW or pGLOW containing the bp -397 of human *Trail* promoter in the presence of the β -galactosidase-expressing vector. Transfected cells were incubated for 13 h before stimulation with GP₃₃₋₄₁ peptide. *GFP* mRNA was analysed by real-time RT-PCR and normalized to *18S* and to *LacZ* mRNA (as an efficiency control) expression. Data are shown as mean $\pm$ s.e.m. ($n = 5-6$ mice per group).

Our results shed new light on the physiological role of TRAIL in immune homeostasis. Early studies showed that TRAIL can induce apoptotic death in a variety of transformed cells while sparing normal cells, leading to an interest in the clinical use of TRAIL for cancer immunotherapy^{7,16}. Of relevance to immune function, TRAIL has been shown to contribute to activation-induced apoptosis of the human leukaemia cell line Jurkat, *in vitro* activated human CD4⁺ T cell clones, and in peripheral blood lymphocytes^{17,18}. Our study demonstrates that TRAIL can regulate peripheral CD8⁺ T-cell responses through CD4⁺ T help, a finding that may be relevant to the establishment and maintenance of peripheral tolerance. Like helpless CD8⁺ T cells, many potentially autoreactive T cells in the periphery will encounter their cognate antigen on a non-professional or immature antigen-presenting cell (APC) lacking sufficient stimulatory activity, thereby potentially inducing a 'helpless' phenotype. Re-encounter with self-antigen would then result in clonal deletion by means of TRAIL-mediated AICD, thereby purging the peripheral repertoire of self-reactive clones. The transwell experiments (Fig. 4c, d) indicate that the TRAIL produced by helpless CD8⁺ T cells is soluble and can therefore potentially act on adjacent cells that are not in direct contact. During a normal response it is possible that not all CD8⁺ T cells receive the help necessary to develop into memory cells, and the emerging population would be a mix of helped and helpless cells. The capacity of helpless CD8⁺ T cells to secrete TRAIL and thereby suppress other cells in their proximity upon re-encounter with antigen may represent a novel mechanism of immune regulation.

In conclusion, by using a range of antigenic challenges, our study reveals that the presence or absence of CD4⁺ T help is functionally imprinted in CD8⁺ T cells at an early time point during their clonal activation as they transit from the naive to the primed state. Depending on the availability of help, CD8⁺ T cells will acquire distinct genetic programmes that each begin with proliferation and effector differentiation, but which specify fundamentally different fates for their clonal progeny upon restimulation (AICD versus secondary expansion). The differential regulation of *Trail*

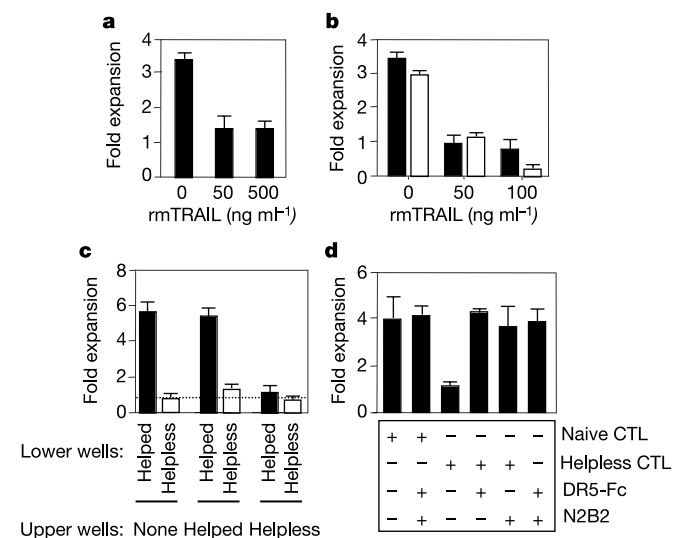


Figure 4 Suicide and fratricide by helpless CD8⁺ T-cell-derived *Trail*. **a**, **b**, Purified CD8⁺ T cells from intact wild-type mice (**a**) and intact and CD4-depleted *Trail*^{-/-} mice (**b**) (helped, filled bar; helpless, open bar) were cultured with Tap^{+/+} Ad5E1-MEC in the presence or absence of recombinant murine (rm)TRAIL for 6 days and the fold expansion of E1B₁₉₂₋₂₀₀-specific CD8⁺ T cells was determined. **c**, Purified helped (filled bar) and helpless (open bar) CD8⁺ T cells were cultured with Tap^{+/+} Ad5E1-MEC in the bottom of transwell plates. In the top chamber, purified helped or helpless CD8⁺ T cells were cultured with Tap^{+/+} Ad5E1-MEC. After 6 days the fold expansion of E1B₁₉₂₋₂₀₀-specific CD8⁺ T cells in the bottom chamber was determined. Horizontal dotted line shows the fold expansion of helpless CD8⁺ T cells. **d**, Helped CD8⁺ T cells (CTL) were cultured with Tap^{+/+} Ad5E1-MEC at the bottom of transwell plates with purified naive or helpless CD8⁺ T cells with Tap^{+/+} Ad5E1-MEC in the top chamber. TRAIL-blocking antibody or soluble DR5-Fc were added for the entire duration of culture. After 6 days the fold expansion of the helped CD8⁺ T cells in the bottom chamber was determined as the increase in the absolute number of E1B₁₉₂₋₂₀₀-specific CD8⁺ T cells. Data are shown as mean $\pm$ s.e.m. ($n = 3$ mice per group).

expression at this later time point clearly represents a key feature of this mechanism and can account for the observed role of CD4⁺ T cells in the generation of CD8⁺ T cell memory. It is interesting to note that helpless CD8⁺ T cells are not irretrievably consigned to death by TRAIL-mediated AICD upon restimulation, however, because cytokines such as IL-2 can permit their re-expansion when provided during secondary antigenic encounter⁴. It is possible that other cytokines may mediate similar effects on the survival of helpless cells¹⁹. The signals through which the molecular signature of T-cell help is transmitted to CD8⁺ T cells during their priming, and the mechanism through which these are integrated in secondary responses remain intriguing questions. □

Methods

Mice and cell lines

C57BL/6J, LPR/LPR, TNFR1^{-/-}, B6.SJL-Ptpcr (B6/SJL) and B6.SJL/I-A^b^{-/-} (all H-2b) mice were purchased from The Jackson Laboratory. Bcl-2-Tg, Bcl-xL-Tg, *Trail*^{-/-} and act-mOVA mice on a C57BL/6J background have been previously described^{20–22}. The act-mOVA/K^b^{-/-} variant was generated by intercrossing with the H-2K^b^{-/-} strain²³. Mice were maintained by in-house breeding at the La Jolla Institute for Allergy and Immunology under specific pathogen-free conditions in accordance with guidelines by the Association for Assessment and Accreditation of Laboratory Animal Care International. Mouse embryo cell lines (MEC) expressing the human adenovirus type 5 early region 1 (Ad5E1) were produced by transfection of both C57BL/6 TAP^{+/+} and TAP^{-/-} MEC lines and have been previously described⁵. MEC were cultured in DMEM supplemented with 10% fetal calf serum, 50 µM 2-mercaptoethanol, 2 mM L-glutamine, 20 U ml⁻¹ penicillin, and 20 µg ml⁻¹ streptomycin.

Immunizations and antibody treatment

Mice were immunized subcutaneously in the right flank with 1 × 10⁷ irradiated (3,000 rad) TAP^{-/-}-Ad5E1-MEC, 2 × 10⁷ irradiated (1,500 rad) act-mOVA/K^b^{-/-} splenocytes, or were inoculated intraperitoneally with 2 × 10⁵ PFU LCMV Armstrong. Depletion of CD4⁺ cells *in vivo* was performed by intraperitoneal administration of 150 µg GK1.5 antibody on the first 3 days before immunization (CD4-depleted at the time of priming) or 3 days after immunization (intact at time of priming)⁴. Administration of GK1.5 antibody was continued every 3 days in all mice for the entire length of the experiments to prevent repopulation²⁴.

Isolation and stimulation of CD8⁺ T cells

CD8⁺ T cells were purified from the spleens and lymph nodes of previously immunized mice by antibody-directed complement lysis or FACS sorting after GP_{33–41}/K^b tetramer staining²⁵. The resulting cells were >95% pure CD8⁺ T cells, and contained less than 0.1% CD4⁺ T cells as demonstrated by FACS analysis. In some expansion studies CD8⁺ T cells were CFSE-labelled before use.

For E1B-specific responses, purified CD8⁺ T cells from TAP^{-/-}-Ad5E1-MEC-immunized mice were stimulated *in vitro* for 6 days with irradiated syngeneic TAP^{+/+}-Ad5E1-MEC (10:1 ratio). CD8⁺ T cells from mice immunized with act-mOVA/K^b^{-/-} splenocytes were stimulated for 6 days *in vitro* at a 10:1 ratio with irradiated MEC.B7.Sig-OVA cells⁴. CD8⁺ T cells from LCMV-infected mice were stimulated for 7 days *in vitro* at a 10:1 ratio with irradiated syngeneic I-A^b^{-/-} thiolglycollate induced macrophages infected with LCMV or pulsed with GP_{33–40} peptide (KAVYNFATC; 5 µg ml⁻¹). Transwell studies were performed in 24-well plates using polycarbonate membrane inserts (0.4-µm pore, Corning Inc.). qVD-OPH (quinoline-Val-Asp-CH₂-difluorophenoxy) and zVAD-fmk (benzyloxycarbonyl-Val-Ala-Asp-fluoromethyl ketone) (20 µM, ICN), TNFR1-Fc, Fas-Fc, TRAIL-R-Fc (all 5 µg ml⁻¹, R&D Systems), or N2B2 (10 µg ml⁻¹, eBioscience) and rmTRAIL (Biomol) were added to indicated cultures. At the end of all stimulations, viable cells were collected by Ficoll gradient (Cedarlane Laboratories).

Enumeration of antigen-specific CD8⁺ T cells

Spleen and lymph node cells were incubated for 5 h with E1B_{192–200} peptide (VNIRNCCYI), OVA_{257–264} (SIINFEKL), or GP_{33–40} (KAVYNFATC) at 5 µg ml⁻¹ final concentration in the presence of Brefeldin A directly *ex vivo*, after CD8 purification, or following *in vitro* culture. Surface staining for CD8 and intracellular cytokine staining for IFN-γ and TNF was performed using a Cytofix/Cytoperm Kit (Pharmingen) according to the manufacturer's directions. The fold expansion of specific CD8⁺ T cells was calculated by dividing the absolute number of IFN-γ⁺ CD8⁺ cells after *in vitro* culture by the absolute number of IFN-γ⁺ CD8⁺ cells at the start of the culture.

Real-time reverse transcription-PCR (RT-PCR)

Purified CD8⁺ T cells were stimulated with their cognate peptides for various lengths of time. CD8⁺ T cells were either used directly or Ag-specific CD8⁺ T cells were further purified by FACS sorting in combination with cytokine staining for IFN-γ or specific peptide/MHC-tetramer. Total RNA was isolated from purified CD8⁺ T cells using TriZol (Gibco BRL) according to the manufacturer's instructions. RNA was reverse transcribed by M-MLV reverse transcriptase (Gibco BRL) using random hexamers (Gibco BRL). Sequence-specific primers for murine *FasL*, *TNF*, *TRAIL*, *TRAIL-R2/DR5*, *GFP* and *LacZ*, and *18S* were previously described^{13,26}. Specific primers for murine *Bcl-2* (forward primer: 5'-ACTTCGAGAGATGTCAGTCA-3'; reverse primer: 5'-TGGCAAAGCGTCCC

CTC-3'); *Bcl-xL* (forward primer: 5'-TCGGGATGGAGTAACTGGG-3'; reverse primer: 5'-CCACGCACAGTCCCC-3'); *FLIP* (forward primer: 5'-AGCAACCGTGGAGGACCA-3'; reverse primer: 5'-CCATCAGCAGGACCCCTATAATCA-3') and *L32* (forward primer: 5'-GAAACTGGCGGAACCCA-3'; reverse primer: 5'-GGATCTGGCCCTTG AACCTT-3') were used. Real-time PCR was performed with AmpliTaq Gold polymerase in a PE Biosystems 5700 thermocycler using SyBr Green detection protocol as outlined by the manufacturer. Briefly, 12 ng of total complementary DNA, 50 nM of each primer and 1X SyBr Green mix were used in a total volume of 25 µl. *L32* and *18S* were used as internal controls.

Transient transfection and reporter assays

In transient expression experiments, all transfections were performed using Superfect (Qiagen) according to the manufacturer's instructions with 0.2 µg of each vector. Helped or helpless CD8⁺ T cells (1.5 × 10⁶) were transfected and incubated 13 h at 37 °C before stimulation *in vitro* with E1B_{192–200} or GP_{33–41} peptides. Owing to the low efficiency of transient transfection, *GFP* mRNA was analysed by real-time PCR and transfection efficiencies were measured by analysis of *LacZ* reporter mRNA expression.

Human *Trail* promoter construct (397 bp) was generated by PCR using the 5'-CGACGCGTCCACATATGAAGTTCAGGTC-3' and the 5'-GGAAGATCTTGAAA GCGAATGAGTGTGTTTTCTGGG-3' primers. Amplified fragment was cloned directly into pGLOW vector (Invitrogen) and sequenced according to the manufacturer's instructions¹³.

Received 25 October; accepted 17 December 2004; doi:10.1038/nature03337.

- Shedlock, D. J. & Shen, H. Requirement for CD4 T cell help in generating functional CD8 T cell memory. *Science* **300**, 337–339 (2003).
- Sun, J. C. & Bevan, M. J. Defective CD8 T cell memory following acute infection without CD4 T cell help. *Science* **300**, 339–342 (2003).
- Bourgeois, C., Rocha, B. & Tanchot, C. A role for CD40 expression on CD8⁺ T cells in the generation of CD8⁺ T cell memory. *Science* **297**, 2060–2063 (2002).
- Janssen, E. M. *et al.* CD4⁺ T cells are required for secondary expansion and memory in CD8⁺ T lymphocytes. *Nature* **421**, 852–856 (2003).
- Schoenberger, S. P. *et al.* Cross-priming of CTL responses *in vivo* does not require antigenic peptides in the endoplasmic reticulum of immunizing cells. *J. Immunol.* **161**, 3808–3812 (1998).
- Lyons, A. B. & Parish, C. R. Determination of lymphocyte division by flow cytometry. *J. Immunol. Methods* **171**, 131–137 (1994).
- Pitti, R. M. *et al.* Induction of apoptosis by Apo-2 ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* **271**, 12687–12690 (1996).
- Wiley, S. R. *et al.* Identification and characterization of a new member of the TNF family that induces apoptosis. *Immunity* **3**, 673–682 (1995).
- Walczak, H. *et al.* TRAIL-R2: a novel apoptosis-mediating receptor for TRAIL. *EMBO J.* **16**, 5386–5397 (1997).
- Pan, G. *et al.* An antagonist decoy receptor and a death domain-containing receptor for TRAIL. *Science* **277**, 815–818 (1997).
- MacFarlane, M. *et al.* Identification and molecular cloning of two novel receptors for the cytotoxic ligand TRAIL. *J. Biol. Chem.* **272**, 25417–25420 (1997).
- Ehst, B. D., Ingulli, E. & Jenkins, M. K. Development of a novel transgenic mouse for the study of interactions between CD4 and CD8 T cells during graft rejection. *Am. J. Transplant.* **3**, 1355–1362 (2003).
- Droin, N. M., Pinkoski, M. J., Dejardin, E. & Green, D. R. Egr family members regulate nonlymphoid expression of Fas ligand, TRAIL, and tumor necrosis factor during immune responses. *Mol. Cell. Biol.* **23**, 7638–7647 (2003).
- Bhardwaj, A. & Aggarwal, B. B. Receptor-mediated choreography of life and death. *J. Clin. Immunol.* **23**, 317–332 (2003).
- Aggarwal, B. B. Signalling pathways of the TNF superfamily: a double-edged sword. *Nature Rev. Immunol.* **3**, 745–756 (2003).
- Smyth, M. J. *et al.* Nature's TRAIL—on a path to cancer immunotherapy. *Immunity* **18**, 1–6 (2003).
- Zhang, X. R. *et al.* Reciprocal expression of TRAIL and CD95L in Th1 and Th2 cells: role of apoptosis in T helper subset differentiation. *Cell Death Differ.* **10**, 203–210 (2003).
- Martinez-Lorenzo, M. J. *et al.* Involvement of APO2 ligand/TRAIL in activation-induced death of Jurkat and human peripheral blood T cells. *Eur. J. Immunol.* **28**, 2714–2725 (1998).
- Sun, J. C., Williams, M. A. & Bevan, M. J. CD4⁺ T cells are required for the maintenance, not programming, of memory CD8⁺ T cells after acute infection. *Nature Immunol.* **5**, 927–933 (2004).
- Cretney, E. *et al.* Increased susceptibility to tumor initiation and metastasis in TNF-related apoptosis-inducing ligand-deficient mice. *J. Immunol.* **168**, 1356–1361 (2002).
- Strasser, A., Harris, A. W. & Cory, S. bcl-2 transgene inhibits T cell death and perturbs thymic self-censorship. *Cell* **67**, 889–899 (1991).
- Grillot, D. A., Merino, R. & Nunez, G. Bcl-XL displays restricted distribution during T cell development and inhibits multiple forms of apoptosis but not clonal deletion in transgenic mice. *J. Exp. Med.* **182**, 1973–1983 (1995).
- Pascolo, S. *et al.* HLA-A2.1-restricted education and cytolytic activity of CD8(+) T lymphocytes from beta2 microglobulin (beta2m) HLA-A2.1 monochain transgenic H-2Db beta2m double knockout mice. *J. Exp. Med.* **185**, 2043–2051 (1997).
- Dialynas, D. P. *et al.* Characterization of the murine T cell surface molecule, designated L3T4, identified by monoclonal antibody GK1.5: similarity of L3T4 to the human Leu-3/T4 molecule. *J. Immunol.* **131**, 2445–2451 (1983).
- van Stipdonk, M. J., Lemmens, E. E. & Schoenberger, S. P. Naive CTLs require a single brief period of antigenic stimulation for clonal expansion and differentiation. *Nature Immunol.* **2**, 423–429 (2001).
- Pinkoski, M. J., Droin, N. M. & Green, D. R. Tumor necrosis factor alpha up-regulates non-lymphoid Fas-ligand following superantigen-induced peripheral lymphocyte activation. *J. Biol. Chem.* **277**, 42380–42385 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank H. Grey for critical evaluation of the manuscript and S. Jameson for providing peptide/MHC tetramers. This work was supported in part by grants from the National Institutes of Health and the American Cancer Society and by a gift from the estate of Wilton N. Chamberlain. This is manuscript number 649 from the La Jolla Institute for Allergy and Immunology.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.P.S. (sps@liai.org) or D.R.G. (doug@liai.org).

Two pathways converge at CED-10 to mediate actin rearrangement and corpse removal in *C. elegans*

Jason M. Kinchen^{1,2*}, Juan Cabello³, Doris Klingele², Kelvin Wong², Richard Feichtinger*, Heinke Schnabel*, Ralf Schnabel³ & Michael O. Hengartner²

¹Department of Molecular Genetics and Microbiology, State University of New York, Stony Brook, New York 11743, USA

²Institute of Molecular Biology, University of Zurich, Winterthurerstrasse 190, CH - 8057 Zurich, Switzerland

³Institut für Genetik, TU Braunschweig, Spielmannstrasse 7, D - 38106 Braunschweig, Germany

* Present addresses: Beirne Carter Center for Immunology Research, University of Virginia, Charlottesville, Virginia 22908, USA (J.M.K.); Xybermind GmbH, Lorettoplatz 26, D - 72072 Tübingen, Germany (R.F.); Max Planck Institut für Neurobiologie, Am Klopferspitz 18A, D - 82152 Martinsried, Germany (H.S.)

The removal of apoptotic cells is essential for the physiological well being of the organism^{1–3}. In *Caenorhabditis elegans*, two conserved, partially redundant genetic pathways regulate this process^{4–6}. In the first pathway, the proteins CED-2, CED-5 and CED-12 (mammalian homologues CrkII, Dock180 and ELMO, respectively) function to activate CED-10 (Rac1)^{7,8}. In the second group, the candidate receptor CED-1 (CD91/LRP/SREC) probably recognizes an unknown ligand on the apoptotic cell⁹ and signals via its cytoplasmic tail to the adaptor protein CED-6 (hCED-6/GULP)^{10,11}, whereas CED-7 (ABCA1) is thought to play a role in membrane dynamics¹². Molecular understanding of how the second pathway promotes engulfment of the apoptotic cell is lacking. Here, we show that CED-1, CED-6 and CED-7 are required for actin reorganization around the apoptotic cell corpse, and that CED-1 and CED-6 colocalize with each other and with actin around the dead cell. Furthermore, we find that the CED-10(Rac) GTPase acts genetically downstream of these proteins to mediate corpse removal, functionally linking the two engulfment pathways and identifying the CED-1, -6 and -7 signalling module as upstream regulators of Rac activation.

The nematode *C. elegans* provides a powerful genetic system for the study of programmed cell death. Large numbers of cells undergo apoptosis during two periods in the life of the worm: 131 somatic cells die during embryonic and larval development¹³, whereas several hundred germ cells in the adult hermaphrodite gonad will die in response to physiological signals, genotoxic stress and/or bacterial infection^{14,15}. The *C. elegans* hermaphrodite gonad is composed of two U-shaped tubes joined together at a central uterus; germ cell nuclei line the periphery of the tube and are connected with each other through a central core of cytoplasm, forming a large syncytium. When cells die, they cellularize away from the common syncytium and begin to condense, which results in the generation of cell corpses that appear as 'refractile' bodies, as visualized by

differential interference contrast (DIC) optics¹⁶ (Fig. 1). During development, many different cell types can phagocytose (or engulf) the condensed apoptotic corpse¹⁷, suggesting that all cells possess the ability (and machinery) to engulf. However, in the hermaphrodite gonad, only the somatic sheath cells, which encase the germ line, have the ability to engulf¹⁶.

Previous experiments have demonstrated that in mammals CrkII, Dock180, ELMO and Rac (the homologues of CED-2, -5, -12 and -10) promote actin cytoskeleton rearrangement during cell migration and phagocytosis⁵. The worm homologues have been proposed to function similarly during corpse removal in *C. elegans*⁷, although this has never been experimentally tested. Furthermore, the molecular readout of CED-1 and CED-6 activity is unknown. CED-1 interacts biochemically with CED-6 (ref. 11), but when this interaction occurs, and for what stage of engulfment it is required for—corpse recognition, tethering of the apoptotic cell to the engulfing cell, a later stage in corpse degradation, or a combination of the three—is uncharacterized.

To further test the relationship of the two pathways to actin reorganization, we fused the cytosolic actin isoform, *act-5*, to yellow fluorescent protein (YFP) and expressed it in the sheath cell (see Methods). In wild-type worms, we could clearly observe YFP::actin 'halos' around early apoptotic corpses (Fig. 1a, b and Table 1a). Additionally, YFP::actin incorporated into filamentous, phalloidin-stained structures, suggesting that the fusion is functional (Fig. 1i–n and Supplementary Fig. 1). *gla-1*(*op234*) mutant worms, which have increased cell death in the hermaphrodite germ line (S. Milstein and M.O.H., unpublished observations), showed increased numbers of both halos and apoptotic cells (Fig. 1c–d and Table 1a). In contrast, no halos were observed in the gonads of worms mutant for the *ced-3* caspase homologue in which apoptosis is blocked (Table 1a), suggesting that actin halos specifically surround apoptotic germ cells.

To determine at which stage of engulfment actin halos form, we stained transgenic worms with Hoechst 33342 (to monitor DNA condensation) and SYTO 41 (which specifically stains engulfed apoptotic cells; Supplementary Table 1, Supplementary Fig. 2e–h and our unpublished observations). Engagement of the CED-1 receptor is required for the generation of TUNEL-positive bodies¹⁸, implying that the engulfment process begins very early in the apoptotic programme, concomitant with induction of DNA degradation. YFP::actin preferentially highlighted cells in the early stages of apoptosis; these early corpses did not stain intensely with either Hoechst 33342 or SYTO 41 (Fig. 1t, arrowhead). Onset of Hoechst 33342 and SYTO 41 staining appears to occur later in the death process, once the cell has been engulfed and the actin network around the engulfed cell has been dismantled (Fig. 1o–t, asterisks).

We next asked which of the previously identified engulfment genes are required for actin reorganization around the apoptotic cell. Actin halos were absent in *ced-2*, *ced-5*, *ced-10* and *ced-12* mutants (Table 1a), consistent with a role for these genes in actin cytoskeletal rearrangement during engulfment^{5,8}. Interestingly, we obtained a similar result with mutants in the second pathway: even though *ced-1*, -6, and -7 mutants contain many unengulfed corpses, there were very few YFP::actin halos (Table 1a and Fig. 1e–h). These results suggest that CED-1-, CED-6- and CED-7-mediated signalling is specifically required (either directly or indirectly) for reorganization of the actin cytoskeleton in the somatic sheath cell following recognition of the apoptotic cell corpse. Other actin structures, such as stress fibre-like filaments, were still observed in all mutant backgrounds (data not shown).

Previous work has shown that CED-1::GFP (green fluorescent protein) clusters around apoptotic cells⁹; this clustering probably occurs before engulfment, because it can still be observed in most engulfment mutant backgrounds. To gain a clearer understanding of CED-1 and CED-6 function during engulfment, we used CED-1::YFP and CFP::CED-6 functional fusion proteins (see

Supplementary Table 2) to address protein localization during corpse removal. CED-1 and CED-6 were only seen to colocalize around early apoptotic cells (Fig. 2a–d), before onset of DNA degradation (data not shown). This suggests that CED-1 and CED-6 are removed from the membrane following phagocytosis

of the dying cell, and is consistent with a role for CED-1 and CED-6 in the recognition of the apoptotic cell. Both CED-1 and CED-6 colocalized with actin halos (Fig. 2e–l), suggesting that these proteins function early in the engulfment process, either before or during actin cytoskeleton rearrangement. Importantly, CED-1 did

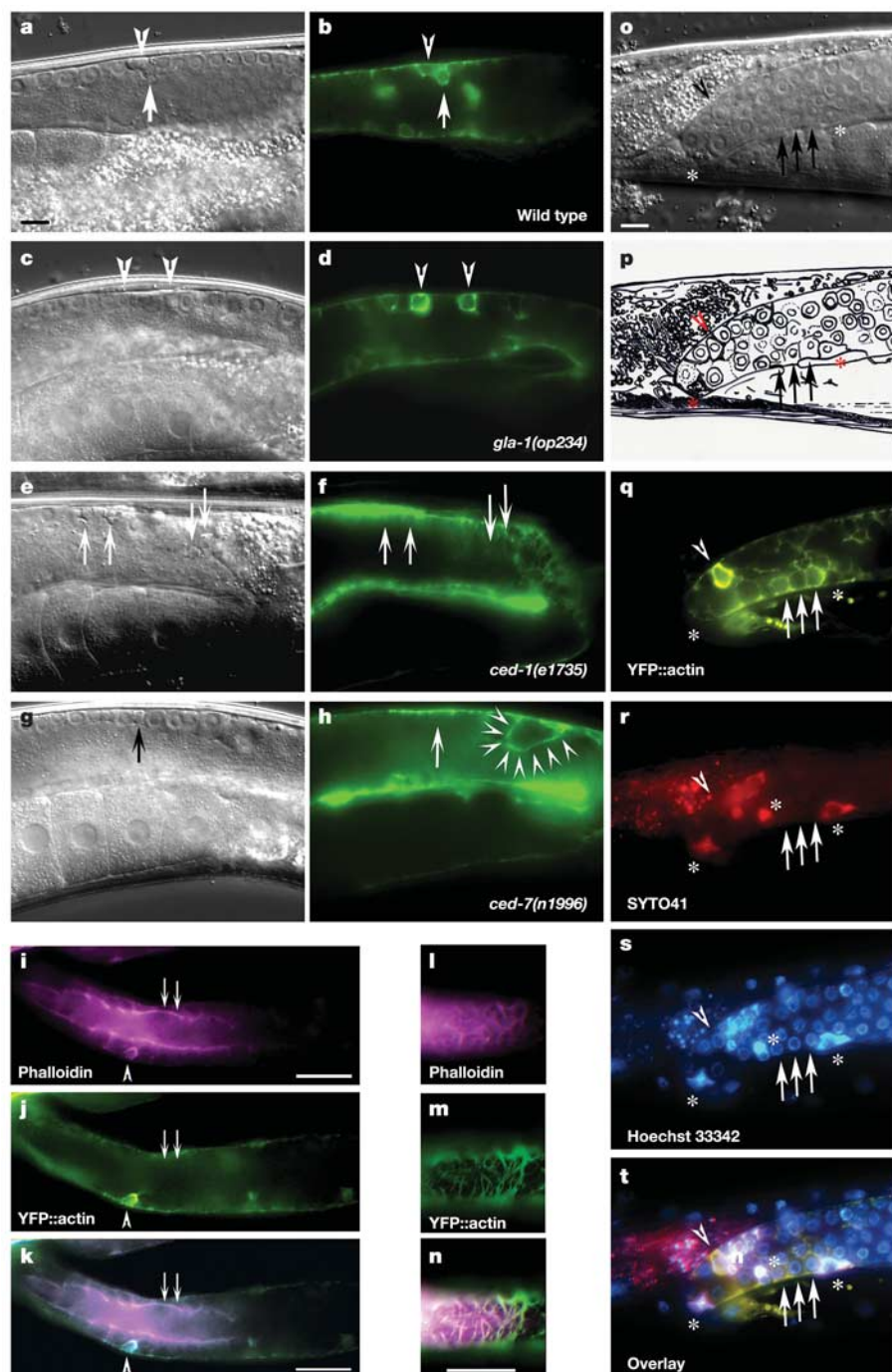


Figure 1 Recruitment of actin halos around apoptotic cells in the hermaphrodite germ line requires CED-1 and CED-7. **a–h**, Arrows/arrowheads indicate refractile corpses (DIC) or YFP::actin halos. In wild-type worms (**a, b**), both refractile corpses (arrowheads) and early stage pre-refractile corpses (arrows) are highlighted; corpse number increases in *gla-1* mutants (**c, d**). Mutations in the engulfment apparatus (*ced-1(e1735)* (**e, f**) and *ced-7(n1996)* (**g, h**)) abrogate actin staining around the apoptotic cell corpse. **f, h**, Pictures were overexposed so intrusions of the somatic sheath between oocytes became visible (**h**, white arrowheads). Scale bar, 10 μ m. **i–n**, Phalloidin staining (**i, l**) colocalized with YFP::actin staining (**j, m**) both around apoptotic cells (**k**, arrowhead (overlay)) and in filamentous actin structures (**n**, (overlay)). F-actin structures within the syncytial germ line

(**l**, arrows) do not colocalize with YFP::actin, which is expressed only in the somatic sheath cells. Scale bars, 40 μm . **o-t**, YFP::actin transgenic animals (**o**, DIC; **p**, camera lucida) were stained with Hoechst 33342 (**s**), which measures condensation of DNA, and with SYTO 41 (**r**), which stains engulfed apoptotic cells. YFP::actin halos highlight cells at an early stage of apoptosis, before they begin to condense their DNA (**q**, arrows; **t**, overlay). As apoptosis progresses, the cell begins to condense and the halo becomes more intense (**q**, arrowhead). Subsequent to this, halos disappear as the cell is engulfed (becomes SYTO 41-positive) and its chromatin strongly condenses (**r**, **s**, asterisks). Note that Hoechst- and SYTO-staining are frequently seen to leach out of the cell corpse nucleus following engulfment. Scale bar, 10 μm .

not colocalize with fibrous actin structures in the proximal gonad (Supplementary Fig. 2), confirming that recruitment of actin to CED-1 requires the presence of an apoptotic cell.

We next looked at the genetic requirements for correct localiza-

tion of CED-6 around the corpse. YFP::CED-6 was not localized properly in *ced-1* mutant worms (Fig. 2q–t and Table 1b), suggesting that modification of CED-1 by ‘recognition’ of the apoptotic cell results in CED-6 recruitment. Surprisingly, mutations

Table 1 **CED-1, -6 and -7 function upstream of actin reorganization around cell corpses**

a *ced-1*, -6 and -7 are required for actin reorganization around the apoptotic cell corpse

Genotype	Refractile corpses (DIC)	YFP::actin halos	<i>n</i>
wild type	3.1 ± 1.1		20
<i>unc-119(ed3)</i>	3.1 ± 1.4		14
<i>unc-119(ed3); opls110 (P_{lin-7::yfp::act-5})</i>	3.8 ± 1.2	3.0 ± 1.6	10
<i>unc-69(e587); opEx1007 (P_{lin-7::yfp::act-5})</i>	2.5 ± 0.7	2.1 ± 1.0	10
<i>ced-3(RNAi)</i>	0.1 ± 0.3		10
<i>ced-3(RNAi); opls110 (P_{lin-7::yfp::act-5})</i>	0.2 ± 0.4	0.2 ± 0.6	10
<i>ced-3(RNAi); opEx1007 (P_{lin-7::yfp::act-5})</i>	0	0.2 ± 0.4	10
<i>gla-1(op234)</i>	13.5 ± 5.6		14
<i>gla-1(op234); opls110 (P_{lin-7::yfp::act-5})</i>	10.0 ± 1.8	8.4 ± 3.1	10
<i>gla-1(op234); opEx1007 (P_{lin-7::yfp::act-5})</i>	9.9 ± 2.8	8.6 ± 1.5	10
<i>ced-1(e1735)</i>	18.3 ± 4.0		11
<i>ced-1(e1735); opls110 (P_{lin-7::yfp::act-5})</i>	18.6 ± 4.7	0	10
<i>ced-1(e1735); opEx1007 (P_{lin-7::yfp::act-5})</i>	18.7 ± 2.7	0.2 ± 0.4	10
<i>ced-6(n1813)</i>	18.1 ± 4.2		20
<i>ced-6(n1813); opls110 (P_{lin-7::yfp::act-5})</i>	17.8 ± 3.8	0.1 ± 0.3	10
<i>ced-7(n1892)</i>	15.2 ± 4.3		10
<i>ced-7(n1892); opls110 (P_{lin-7::yfp::act-5})</i>	17.6 ± 3.1	1.0 ± 1.2	10
<i>ced-2(e1752)</i>	13.1 ± 4.1		10
<i>ced-2(e1752) opls110 (P_{lin-7::yfp::act-5})</i>	12.6 ± 2.1	1.3 ± 1.1	10
<i>ced-5(n1812)</i>	15.2 ± 3.6		10
<i>ced-5(n1812) opls110 (P_{lin-7::yfp::act-5})</i>	15.8 ± 4.3	0.3 ± 0.5	10
<i>ced-5(n1812); opEx1007 (P_{lin-7::yfp::act-5})</i>	18.3 ± 3.3	0.2 ± 0.4	10
<i>ced-10(n3246)</i>	24.5 ± 5.6		20
<i>ced-10(n3246); opEx1007 (P_{lin-7::yfp::act-5})</i>	26.9 ± 2.7	0.1 ± 0.3	10
<i>ced-12(k149)</i>	12.6 ± 5.2		11
<i>ced-12(k149); opls110 (P_{lin-7::yfp::act-5})</i>	12.1 ± 2.8	1.0 ± 1.2	10
<i>mig-2(mu28)</i>	3.6 ± 1.4		10
<i>mig-2(mu28); opls110 (P_{lin-7::yfp::act-5})</i>	2.9 ± 0.9	2.0 ± 1.2	10

b CED-1, but not CED-7, is required for localization of CED-6 around apoptotic cell corpses

Genotype	Refractile corpses (DIC)	YFP::CED-6 halos	<i>n</i>
<i>unc-119(ed3); opls160 (P_{ced-6::yfp::ced-6})</i>	2.9 ± 0.3	3.3 ± 1.1	10
<i>ced-1(e1735); opls160 (P_{ced-6::yfp::ced-6})</i>	20.7 ± 6.5	3.8 ± 3.3	10
<i>ced-7(n1892); opls160 (P_{ced-6::yfp::ced-6})</i>	17.4 ± 4.4	13.5 ± 4.9	10
<i>ced-12(k149); opls160 (P_{ced-6::yfp::ced-6})</i>	15.8 ± 2.2	7.2 ± 4.0	10

c Overexpression of CED-10 promotes bypass of the *ced-1*, -6 and -7 engulfment defect

Genotype	Refractile corpses (DIC)	<i>n</i>
Wild type	0	20
<i>opEx700 (hs::ced-10(WT))</i>	0.2 ± 0.4	10
<i>ced-1(e1735)</i>	20.8 ± 6.5	20
<i>ced-1(e1735); opEx700 (hs::ced-10(WT))</i>	7.8 ± 4.5	20
<i>ced-6(n1813)</i>	12.8 ± 6.6	26
<i>ced-6(n1813); opEx700 (hs::ced-10(WT))</i>	2.8 ± 3.3	36
<i>ced-6(n2095)</i>	14.7 ± 8.4	28
<i>ced-6(n2095); opEx700 (hs::ced-10(WT))</i>	4.9 ± 3.8	29
<i>ced-7(n2690)</i>	12.0 ± 4.5	16
<i>ced-7(n2690); opEx700 (hs::ced-10(WT))</i>	4.0 ± 2.7	21
<i>ced-7(n1892)</i>	22.7 ± 6.6	20
<i>ced-7(n1892); opEx700 (hs::ced-10(WT))</i>	6.0 ± 5.6	25
<i>ced-2(e1752)</i>	9.9 ± 5.5	14
<i>ced-2(e1752); opEx700 (hs::ced-10(WT))</i>	0.8 ± 0.9	27
<i>ced-5(n1812)</i>	22.0 ± 7.8	20
<i>ced-5(n1812); opEx700 (hs::ced-10(WT))</i>	6.1 ± 6.8	20
<i>ced-10(n1993)</i>	15.1 ± 6.8	15
<i>ced-10(n1993); opEx700 (hs::ced-10(WT))</i>	0.7 ± 2.2	33
<i>ced-12(k149)</i>	17.3 ± 4.5	15
<i>ced-12(k149); opEx700 (hs::ced-10(WT))</i>	1.2 ± 1.4	20
<i>ced-6(n1813); ced-5(n1812)</i>	35.2 ± 11.4	19
<i>ced-6(n1813); ced-5(n1812); opEx700 (hs::ced-10(WT))</i>	26.4 ± 9.4	41
<i>ced-1(e1735); ced-5(n1812)</i>	40.4 ± 4.1	10
<i>ced-1(e1735); ced-5(n1812); opEx700 (hs::ced-10(WT))</i>	35.2 ± 5.1	30
<i>ced-7(n1892); ced-5(n1812)</i>	37.3 ± 5.1	10
<i>ced-7(n1892); ced-5(n1812); opEx700 (hs::ced-10(WT))</i>	37.5 ± 4.5	15
<i>ced-1(e1735); ced-5(n1812)/+</i>	27.7 ± 6.7	10
<i>ced-1(e1735); ced-5(n1812)/+; opEx700 (hs::ced-10(WT))</i>	11.2 ± 8.3	13
<i>ced-7(n1892); ced-5(n1812)/+</i>	20.2 ± 8.4	12
<i>ced-7(n1892); ced-5(n1812)/+; opEx700 (hs::ced-10(WT))</i>	4.3 ± 3.5	12

a, YFP::actin halos were scored 12 h post-L4/adult molt as described in the Methods. *opls110* also carries a wild-type copy of the co-transformation marker *unc-119*. *opEx1007* carries a wild-type copy of *unc-69*. *n*, number of worms scored; data are shown as average ± s.d. **b**, CED-6 halos were scored 12 h post-L4/adult molt. *unc-119(+)* was used as a co-transformation marker. *n*, number of worms scored; data are shown as average ± s.d. **c**, CED-10 expression was induced and cell corpses scored in the L1 head as described in the Methods. *opEx700* also carries pTG96 (*P_{sur-5::gfp}*) as co-transformation marker. *n*, number of worms scored; data are shown as average ± s.d.

in *ced-7* (or *ced-12*) had little effect on CED-6 localization (Fig. 2u–v and Table 1b). Previous studies have suggested that CED-7 is required for CED-1 recruitment around the corpse. However, in the gonad we find that CED-1::YFP (and CFP::CED-6) is properly recruited (Fig. 2x versus Fig. 2z and Table 1b), suggesting that CED-7 may have alternative roles during removal of somatic and germ cell corpses.

Of the three *rac* homologues in *C. elegans*, only *ced-10* is required for the engulfment of apoptotic cells¹⁹, raising the possibility that CED-10 might mediate actin remodelling downstream of the CED-1, -6, -7 group. Overexpression of CED-10 has previously been used to suggest that *ced-10* functions downstream of *ced-2*, *ced-5* and *ced-12* (refs 5, 20). We thus tested the effect of CED-10 overexpression on corpse persistence in *ced-1*, *ced-6* and *ced-7* mutants. To our surprise, all engulfment mutants showed a reduction in cell corpse number following *ced-10* induction

(Table 1c), suggesting that *ced-10* not only functions downstream of *ced-2*, -5 and -12, but also downstream of (or in parallel to) *ced-1*, -6 and -7 during engulfment of apoptotic cells. Double mutants blocked in both engulfment pathways (for example, *ced-1; ced-5* or *ced-7; ced-5*) showed little or no reduction in persistent cell corpses following induction of CED-10 expression (Table 1c). *ced-6; ced-5* double mutant worms did show some reduction of corpse number following *ced-10* induction (Table 1c), probably due to residual activity of the temperature-sensitive *n1813* allele²¹. This suggests that rescue of the engulfment defect in single mutants by CED-10 is dependent on intact signalling by the parallel engulfment pathway. Expression of CED-10 in a wild-type background or *ced-3(lf)* background did not generate any additional halos (Supplementary Table 3), further confirming the requirement of signals from the apoptotic cell for the generation of actin halos.

The original double-mutant analysis that ordered the engulfment

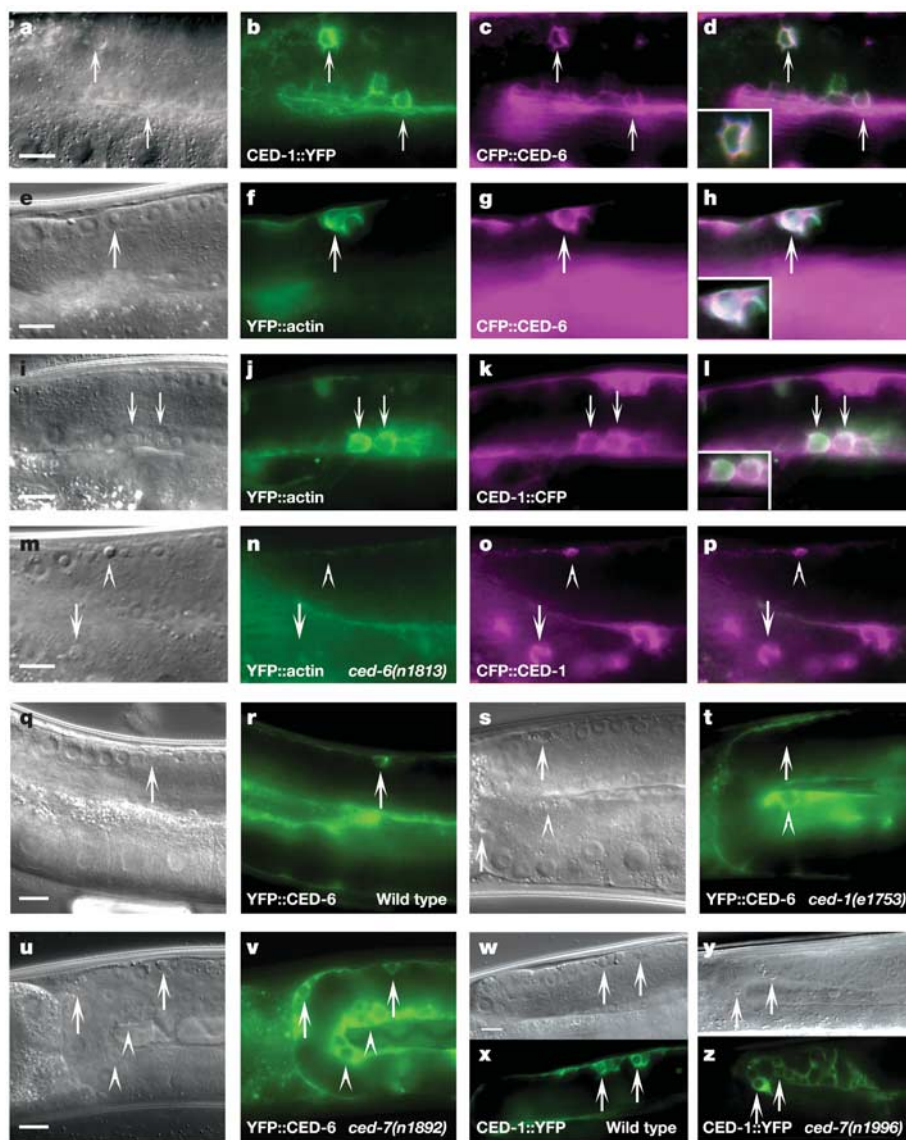


Figure 2 CED-1 recruits CED-6 and actin around the apoptotic cell. Fluorescent marker is annotated in the picture. Scale bar, 10 μ m. **a–d**, CED-1::YFP (**b**) localizes around an apoptotic cell corpse (arrows) in the hermaphrodite gonad (**a**) as previously described⁹. CFP-tagged CED-6 colocalizes with CED-1 (**c**, **d** (overlay), inset). **e–p**, Both CED-1 (**k**, **l** (overlay), inset) and CED-6 (**g**, **h** (overlay), inset) colocalize with actin around apoptotic cells (arrows). In a *ced-6(n1813)* mutant background, CED-1 still localizes around the apoptotic cell (**o**) but no actin reorganization is seen (**n**, **p** (overlay)). Arrowhead denotes partial recruitment of CED-1 staining around the corpse without extension of the

membrane around the apoptotic cell. **q–z**, In wild-type worms, YFP::CED-6 halos apoptotic cells (**r**); in *ced-1(e1735)* mutants, the ability of CED-6 to be recruited around the apoptotic cell is compromised (**t**). Surprisingly, in *ced-7(n1892)* mutants CED-6 recruitment is unperturbed (**v**); similarly, recruitment of CED-1 around the apoptotic cell was unaffected (**w**, **x**, wild-type versus **y**, **z**, *ced-1(e1735)*). Occasionally, persistent unengulfed corpses produce a nonspecific physical distortion of the sheath cell, resulting in circular areas of reduced background staining (**t**, **v**, arrowheads).

genes into a minimum of two gene groups used the weak *ced-10(n1993)* allele⁶ (Table 2a, b), which, due to mutation of the carboxy-terminal CAAX box (Supplementary Fig. 4), probably affects the subcellular localization of CED-10 rather than altering activity²⁰. The use of hypomorphic (weak) alleles in epistatic and double-mutant analyses is problematic, because the requirement for function in one process may be affected, whereas another process may proceed largely unhindered. Indeed, this effect has previously been observed between *ced-12(tp2)* and *ced-10(n1993)*, resulting in inappropriate classification of these two genes into separate pathways²². Subsequent genetic screens have led to the isolation of stronger *ced-10* mutations, such as *ced-10(n3246)*, which is predicted to have a reduced ability to bind GTP²⁰. Here, we have isolated the mutation *ced-10(t1875)* (Supplementary Fig. 3), which results in a lesion that alters the initial ATG (translation start site) of CED-10, and is predicted to be a null allele; a deletion allele,

ced-10(n3417), has also been described¹⁹ (Supplementary Fig. 3).

The *t1875* and *n3417* alleles result in significant embryonic lethality due to defects in ventral enclosure and other cell migrations^{19,23} (data not shown). This embryonic lethality is maternally rescued: *ced-10* mutant embryos generated by *ced-10(+)* mothers survive and develop normally, due to incorporation of *ced-10(+)* gene product into the developing oocytes. It is thus not possible to measure the engulfment defect of *t1875* or *n3417* through the classical larval stage 1 (L1) head corpse assay²⁴, as the first generation of *ced-10* mutants is maternally rescued and the second generation dies before hatching. We instead assessed the effect of these alleles on germ cell corpse removal by scoring the number of persistent cell corpses in the germ line of first generation mutant hermaphrodites, because the engulfment defect in the first generation adult hermaphrodite gonad is not maternally rescued (Table 2b, *n3246* versus *n3246(m + z -)*). As previously described for embryonic cell deaths, *ced-1*; *ced-5* double mutants have an additive effect on germ line corpse number as compared to single mutants, consistent with two pathways mediating engulfment of apoptotic cells⁶ (Table 2). A similar additivity was also seen in double mutants between *ced-1* and the weak CAAX mutant *ced-10(n1993)*, consistent with the previous suggestion that *ced-1* and *ced-10* may act in separate pathways.

Using null alleles of *ced-10* and the other engulfment mutants (Supplementary Table 4), we reassessed whether CED-1, -6 and -7 might function upstream of CED-10 during corpse removal. Interestingly, corpse numbers in *ced-10(t1875)* and *ced-10(n3417)* first-generation adults are similar to numbers seen in *ced-1*; *ced-5* and *ced-1*; *ced-10(n1993)* double mutants, suggesting that *ced-10* may be involved in signalling from both pathways. Indeed, scores taken in double mutants between *ced-10(t1875)* or *ced-10(n3417)* and *ced-1*, *ced-6*, or *ced-7* show no increase in corpse number over the *ced-10* single mutants. Using the strong but viable *ced-10(n3246)* mutation, we obtained a similar pattern for phagocytosis of developmental cell corpses, as scored in the L1 head (Table 2a). We observed a weak genetic interaction between *ced-10(n3246)* and members of the *ced-1*, -6, -7 gene group, possibly due to residual activity present in the non-null *n3246* mutation. These results suggest that both engulfment pathways ultimately converge at CED-10 to facilitate corpse removal in the worm. A fraction of cell deaths are still engulfed even in *ced-10* null worms (albeit with delayed kinetics) (Supplementary Table 5); the underlying mechanism that accomplishes this feat is presently unknown.

Taken together, the data presented here provide important new insights into the mechanism of engulfment in *C. elegans*. Whereas the CED-2, -5, -12 pathway is subject to intense research, before this study little was known of the role of CED-1, -6, or -7 in corpse removal. We show here that both engulfment pathways lead to actin-dependent cytoskeletal reorganization and subsequent engulfment of the apoptotic cell. In addition, we find that CED-1 and CED-6 are recruited around the apoptotic cell either before or concomitantly with actin rearrangement, supporting a role for these proteins in early signalling events. CED-1 is able to cluster around apoptotic cells in the absence of CED-6; however, increased actin staining is not observed (Fig. 2m–p and Table 1a). CED-1 is also required for CED-6 recruitment to the membrane (Fig. 2s, t and Table 1b). Because both CED-1 and CED-6 are required for actin reorganization around the apoptotic cell (see Table 1a), we conclude that, temporally, actin reorganization must occur after CED-1 and CED-6 recruitment around the corpse.

Most importantly, we have identified CED-10/Rac as a downstream mediator of CED-1, -6, -7 signalling, functionally linking the two engulfment pathways. How Rac activation is accomplished by this signalling module remains to be determined. Because the components of the engulfment machinery are highly conserved from worm to human, we believe our observations also have important implications for engulfment in mammals and the

Table 2 Double-mutant analysis reveals a role for CED-10 in both engulfment pathways

Genotype	Refractile corpses (DIC)	n
a Embryonic cell deaths		
wild type	0.1 ± 0.3	10
<i>ced-1(e1735)</i>	15.4 ± 5.1	20
<i>ced-6(n1813)</i>	15.9 ± 3.5	20
<i>ced-7(n1996)</i>	22.8 ± 5.7	20
<i>ced-5(n1812)</i>	21.0 ± 6.3	20
<i>ced-10(n1993)</i>	15.1 ± 4.3	20
<i>ced-10(n3246)</i>	28.5 ± 4.9	20
<i>ced-1(e1735); ced-5(n1812)</i>	35.4 ± 5.1	20
<i>ced-6(n1813); ced-5(n1812)</i>	34.6 ± 4.6	20
<i>ced-7(n1996); ced-5(n1812)</i>	39.2 ± 5.8	20
<i>ced-1(e1735); ced-10(n1993)</i>	27.5 ± 4.3	20
<i>ced-6(n1813); ced-10(n1993)</i>	20.5 ± 2.2	20
<i>ced-7(n1996); ced-10(n1993)</i>	33.0 ± 5.5	20
<i>ced-10(n1993) ced-5(n1812)</i>	22.2 ± 6.4	20
<i>ced-1(e1735); ced-10(n3246)</i>	30.6 ± 4.8	20
<i>ced-6(n1813); ced-10(n3246)</i>	30.8 ± 5.6	20
<i>ced-7(n1996); ced-10(n3246)</i>	34.8 ± 3.0	20
<i>ced-10(n3246) ced-5(n1812)</i>	27.5 ± 4.7	20
b Germ cell deaths		
wild type	3.1 ± 1.1	20
<i>gla-1(op234)</i>	13.5 ± 5.6	14
<i>ced-3(n171)</i>	0.1 ± 0.3	20
<i>ced-1(e1735)</i>	18.2 ± 4.4	20
<i>ced-1(n691)</i>	17.3 ± 2.9	20
<i>ced-5(n1812)</i>	14.7 ± 4.0	20
<i>ced-6(n1813)</i>	18.1 ± 4.2	20
<i>ced-7(n1996)</i>	14.1 ± 3.6	20
<i>ced-7(n1892)</i>	14.4 ± 2.7	20
<i>ced-12(k149)</i>	17.6 ± 3.1	20
<i>ced-10(n1993)</i>	17.5 ± 6.0	20
<i>ced-10(n3246)</i>	24.5 ± 5.6	20
<i>ced-10(n3246, m + z -)</i>	26.2 ± 4.5	20
<i>ced-10(n3417, m + z -)</i>	28.7 ± 5.2	20
<i>ced-10(t1875, m + z -)</i>	32.7 ± 8.3	20
<i>ced-1(e1735); ced-5(n1812)</i>	28.8 ± 7.4	20
<i>ced-6(n1813); ced-5(n1812)</i>	33.1 ± 6.9	20
<i>ced-7(n1996); ced-5(n1812)</i>	32.4 ± 7.2	20
<i>ced-12(k149); ced-7(n1996)</i>	29.8 ± 4.3	20
<i>ced-1(e1735); ced-10(n1993)</i>	26.6 ± 7.1	20
<i>ced-6(n1813); ced-10(n1993)</i>	23.8 ± 4.8	20
<i>ced-7(n1996); ced-10(n1993)</i>	27.7 ± 7.3	20
<i>ced-10(n1993); ced-5(n1812)</i>	18.2 ± 4.7	20
<i>ced-1(e1735); ced-10(n3246)</i>	27.1 ± 6.0	20
<i>ced-6(n1813); ced-10(n3246)</i>	28.1 ± 4.1	20
<i>ced-7(n1996); ced-10(n3246)</i>	28.6 ± 5.4	20
<i>ced-10(n3246); ced-5(n1812)</i>	25.5 ± 5.8	20
<i>ced-12(k149); ced-10(n3246)</i>	23.5 ± 4.5	20
<i>ced-1(n691); ced-10(n3417, m + z -)</i>	29.1 ± 7.3	20
<i>ced-7(n1892); ced-10(n3417, m + z -)</i>	30.7 ± 6.1	20
<i>ced-1(e1735); ced-10(t1875, m + z -)</i>	32.6 ± 6.8	20
<i>ced-1(n691); ced-10(t1875, m + z -)</i>	32.8 ± 7.5	20
<i>ced-6(n1813); ced-10(t1875, m + z -)</i>	33.6 ± 6.8	20
<i>ced-7(n1996); ced-10(t1875, m + z -)</i>	32.7 ± 7.1	20
<i>ced-7(n1892); ced-10(t1875, m + z -)</i>	34.9 ± 7.5	8
<i>ced-12(k149); ced-10(t1875, m + z -)</i>	31.5 ± 7.5	20

Single and double mutants were scored for persistent developmental cell corpses (a) and germ cell corpses (b) as described in the Methods. *m + z -*, homozygous mutant progeny of heterozygous mothers; in all other cases, homozygous mutant progeny from homozygous mutant mothers (*m - z -*) were scored. *n*, number of worms scored; data are shown as average ± s.d.

identification/characterization of functional homologues of CED-1 will be of great scientific interest. CED-1 is a member of the scavenger receptor superfamily; the most likely CED-1 homologue, LRP, is involved in diverse processes, acting both as a signalling molecule during neuronal cell migration (via DAB-1) and in MAPK signalling (via JIP-1/2) (ref. 25). In addition, LRP has been shown to have roles in human diseases such as Alzheimer's and atherosclerosis²⁵. The role of Rac in these LRP-dependent processes remains to be determined. □

Methods

C. elegans strain maintenance

Worms were cultured at 20 °C as described²⁶. Mutations used were as follows: linkage group I (LG I): *glu-1*(*op234*) (S. Milstein and M.O.H., unpublished observations), *ced-12*(*k149*), *ced-1*(*e1735*), *ced-1*(*n691*); LG III: *ced-6*(*n1813*), *ced-6*(*n2095*), *ced-7*(*n1892*), *ced-7*(*n1996*), *ced-7*(*n2690*); LG IV: *ced-2*(*e1752*), *ced-10*(*n1993*), *ced-10*(*n3246*), *ced-10*(*t1875*) (this study), *dpy-20*(*e1282ts*); *unc-24*(*e138*), *ced-5*(*e1812*), *him-8*(*e1489*), *opIs110* (*P_{lim-7}::yfp::act-5*) (this study); LG V: *dpy-11*(*e224*), *him-5*(*e1490*). *opEx700* (*hs::ced-10*) (ref. 6) and *opEx1007* (*yfp::actin*; *unc-69*(+)) (this study) are extrachromosomal arrays. The *nT1 qIs51* IV; V translocation was used to balance *t1875* and *n3417*. Unless noted otherwise, mutations were previously described²⁷. *opIs110* appears to be integrated on LG IV near the *ced-10* locus, but was not mapped further; *opIs159* (*ced-1::cfp*; *unc-69*(+)), *opIs160* (*yfp::ced-6*; *unc-119*(+)) and *opIs161* (*cfp::ced-6*; *unc-119*(+)) were not mapped.

To inactivate *ced-3* by RNA interference, worms were synchronized as L1s by hypochlorite treatment; following this, worms were transferred to NGM agarose plates seeded with HT115 (DE3) bacteria that had been transformed with plasmid *L4440LB*, which carries a fragment of *ced-3* cloned between two inverted T7 promoters (for generating double-stranded RNA).

DIC and immunofluorescence microscopy

Worms were mounted on 4% agarose pads in M9, anaesthetized with 3–5 mM levamisole (adults) or 10 mM NaN₃ (embryos), then sandwiched under a coverslip for observation on a Leica DMRA microscope equipped with standard epifluorescence and appropriate filter sets for visualizing YFP, GFP, Hoechst 33342, or CFP/SYTO 41. Images were taken using an Orca-ER camera using OpenLab software, which was also used for false colouring and editing along with Adobe Photoshop 7.0.

To visualize condensed, apoptotic DNA, as well as engulfed cells, worms were stained with 2 mg ml⁻¹ Hoechst 33342 (Molecular Probes), and 50 µM solution of SYTO 41 (Molecular Probes) for 1–2 h at room temperature in distilled H₂O with some bacteria, then transferred to a seeded NGM agar plate to de-stain for ~1 h, to decrease background staining in the intestine. For staining with acridine orange (Molecular Probes), worms were incubated with 0.14 mg ml⁻¹ acridine orange in M9 buffer for 1–1.5 h, then observed under epifluorescence.

To visualize F-actin in the somatic sheath cells using phalloidin, worm gonads were dissected in 10 mM levamisole in egg salts on poly-L-lysine coated slides. Gonads were fixed (1.5% formaldehyde, 80 mM NaCl, 20 mM KCl, 10 mM MgCl₂, 5 mM HEPES, pH 7.2 supplemented with 0.5 U Alexa 594 conjugated phalloidin (Molecular Probes) and 0.5 µg ml⁻¹ lysocleithin (Sigma) for 20 min, then washed with PBS + 1% BSA (Sigma) and mounted under a coverslip sealed with nail polish (Astor Wonderlast).

Synchronization and heat-shock rescue

Amount of persisting corpses in the L1 head is subject to some controversy, with scores varying by publication (and laboratory): either 42.8 ± 4.5 (ref. 22), 31.0 ± 3.6 (ref. 19), or 22.5 ± 1.5 (ref. 5) apoptotic cell corpses persisted in the *ced-5*(*n1812*) allele. We thus chose to synchronize larvae as previously described to allow easier comparison of scores². Briefly, gravid adult hermaphrodites were allowed to lay embryos for ~3 h; to induce *ced-10* expression, embryos were then subjected to heat shock in a 33 °C water bath for 30 min, then returned to 20 °C to recover where appropriate. Worms were scored 12 h later as 4-cell gonad L1s. Scores are consistent with those determined by this method previously². To rule out silencing of the transgene in double mutants, *ced-1*; *ced-5* or *ced-7*; *ced-5* animals were outcrossed to *ced-1* or *ced-7* single mutants (respectively). Heterozygous *ced-1*; *ced-5*/+ or *ced-7*; *ced-5*/+ strains showed reduced numbers of corpses (Table 1c) following CED-10 induction, suggesting the transgene retains functionality.

To score cell corpses in the hermaphrodite germ line, clean worms were synchronized by picking hermaphrodites at the L4 larval stage. These worms were allowed to grow for 24 h at 20 °C, then scored in the germ line for persistent corpses and YFP::actin halos where appropriate. Members of the *ced-2*, *ced-5*, *ced-10* and *ced-12* mutant group were chosen for scoring that had as close to wild-type gonad morphology as possible.

Construction of transgenic animals

Low copy transgenic strains were constructed as previously described²⁸. Briefly, test constructs (*P_{lim-7}::yfp::act-5*, *P_{ced-6}::yfp::ced-6*, *P_{ced-6}::cfp::ced-6*) were precipitated with pDPM0016 (*unc-119*(+)) (gift of J. Austin) onto gold beads in an approximately 4:1 ratio, then shot at *unc-119*(*ed3*) mutant worms using a Biolistic PDS-1000 bombardment apparatus (Bio-Rad). Worms were then washed onto multiple seeded large plates and allowed to starve, then chunked to new plates. *unc-119*(+) worms that expressed both constructs were kept and analysed. For the *yfp::actin* construct, two integrants—*opIs109* and *opIs110*—were isolated, both of which had halo apoptotic cells in the germ line.

Subsequent analyses were performed with *opIs110*. *opIs160* (*yfp::ced-6*; *unc-119*(+)) and *opIs161* (*cfp::ced-6*; *unc-119*(+)) were isolated in a similar manner. *opEx1031* (*ced-1::yfp*; *cfp::ced-6*; *rol-6*(*su1006*)) and *opEx1007* (*yfp::actin*; *unc-69*(+)) were created as previously described²⁹; *ced-1* and *ced-6* constructs were injected at 50 ng µl⁻¹ each, with *rol-6*(*su1006*), at 100 ng µl⁻¹, into *ced-1*; *lon-1* *ced-6* mutants to create *opEx1031*, while *P_{lim-7}::yfp::actin* and *unc-69*(+) (M.O.H., unpublished reagent) were each injected at a concentration of 50 ng µl⁻¹ to generate *opEx1007*.

Although there are no *act-5* loss-of-function mutants with which we could have tested our construct for functionality, we could observe incorporation of YFP::actin into phalloidin-positive structures (Fig. 1i–n, Supplementary Figs 1 and 2a–d) similarly to what has been previously described³⁰, suggesting that the fusion protein is functional. All other transgenes were found to rescue mutations in their respective genetic backgrounds (Supplementary Table 2).

Received 18 October; accepted 5 December 2004; doi:10.1038/nature03263.

- Savill, J. & Fadok, V. Corpse clearance defines the meaning of cell death. *Nature* **407**, 784–788 (2000).
- Scott, R. S. *et al.* Phagocytosis and clearance of apoptotic cells is mediated by MER. *Nature* **411**, 207–211 (2001).
- Hanayama, R. *et al.* Autoimmune disease and impaired uptake of apoptotic cells in MFG-E8-deficient mice. *Science* **304**, 1147–1150 (2004).
- Chung, S., Gumienny, T. L., Hengartner, M. O. & Driscoll, M. A common set of genes mediate removal of both apoptotic and necrotic cell corpses in *C. elegans*. *Nature Cell Biol.* **2**, 931–937 (2000).
- Gumienny, T. L. *et al.* CED-12/ELMO, a novel member of the CrkII/Dock180/Rac pathway, is required for phagocytosis and cell migration. *Cell* **107**, 27–41 (2001).
- Ellis, R. E., Jacobson, D. M. & Horvitz, H. R. Genes required for the engulfment of cell corpses during programmed cell death in *Caenorhabditis elegans*. *Genetics* **129**, 79–94 (1991).
- Conradt, B. Cell engulfment, no sooner *ced* than done. *Dev. Cell* **1**, 445–447 (2001).
- Brugnera, E. *et al.* Unconventional Rac-GEF activity is mediated through the Dock180-ELMO complex. *Nature Cell Biol.* **4**, 574–582 (2002).
- Zhou, Z., Hartwig, E. & Horvitz, H. R. CED-1 is a transmembrane receptor that mediates cell corpse engulfment in *C. elegans*. *Cell* **104**, 43–56 (2001).
- Liu, Q. A. & Hengartner, M. O. Candidate adaptor protein CED-6 promotes the engulfment of apoptotic cells in *C. elegans*. *Cell* **93**, 961–972 (1998).
- Su, H. P. *et al.* Interaction of CED-6/GULP, an adapter protein involved in engulfment of apoptotic cells with CED-1 and CD91/low density lipoprotein receptor-related protein (LRP). *J. Biol. Chem.* **277**, 11772–11779 (2002).
- Hamon, Y., Chambenoit, O. & Chimini, G. ABCA1 and the engulfment of apoptotic cells. *Biochim. Biophys. Acta* **1585**, 64–71 (2002).
- Metzstein, M. M., Stanfield, G. M. & Horvitz, H. R. Genetics of programmed cell death in *C. elegans*: past, present and future. *Trends Genet.* **14**, 410–416 (1998).
- Stergiou, L. & Hengartner, M. O. Death and more: DNA damage response pathways in the nematode *C. elegans*. *Cell Death Differ.* **11**, 21–28 (2004).
- Aballay, A. & Ausubel, F. M. Programmed cell death mediated by *ced-3* and *ced-4* protects *Caenorhabditis elegans* from *Salmonella typhimurium*-mediated killing. *Proc. Natl Acad. Sci. USA* **98**, 2735–2739 (2001).
- Gumienny, T. L., Lambie, E., Hartwig, E., Horvitz, H. R. & Hengartner, M. O. Genetic control of programmed cell death in the *Caenorhabditis elegans* hermaphrodite germline. *Development* **126**, 1011–1022 (1999).
- Hoepfner, D. J., Hengartner, M. O. & Schnabel, R. Engulfment genes cooperate with *ced-3* to promote cell death in *Caenorhabditis elegans*. *Nature* **412**, 202–206 (2001).
- Wu, Y. C., Stanfield, G. M. & Horvitz, H. R. NUC-1, a *Caenorhabditis elegans* DNase II homolog, functions in an intermediate step of DNA degradation during apoptosis. *Genes Dev.* **14**, 536–548 (2000).
- Lundquist, E. A., Reddien, P. W., Hartwig, E., Horvitz, H. R. & Bargmann, C. I. Three *C. elegans* Rac proteins and several alternative Rac regulators control axon guidance, cell migration and apoptotic cell phagocytosis. *Development* **128**, 4475–4488 (2001).
- Reddien, P. W. & Horvitz, H. R. CED-2/CrkII and CED-10/Rac control phagocytosis and cell migration in *Caenorhabditis elegans*. *Nature Cell Biol.* **2**, 131–136 (2000).
- Liu, Q. A. & Hengartner, M. O. Human CED-6 encodes a functional homologue of the *Caenorhabditis elegans* engulfment protein CED-6. *Curr. Biol.* **9**, 1347–1350 (1999).
- Wu, Y. C., Tsai, M. C., Cheng, L. C., Chou, C. J. & Weng, N. Y. *C. elegans* CED-12 acts in the conserved crkII/DOCK180/Rac pathway to control cell migration and cell corpse engulfment. *Dev. Cell* **1**, 491–502 (2001).
- Soto, M. C. *et al.* The GEX-2 and GEX-3 proteins are required for tissue morphogenesis and cell migrations in *C. elegans*. *Genes Dev.* **16**, 620–632 (2002).
- Hengartner, M. O., Ellis, R. E. & Horvitz, H. R. *C. elegans* gene *ced-9* protects cells from programmed cell death. *Nature* **356**, 494–499 (1992).
- Herz, J. & Strickland, D. K. LRP: a multifunctional scavenger and signaling receptor. *J. Clin. Invest.* **108**, 779–784 (2001).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Harris, T. W. *et al.* WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res.* **31**, 133–137 (2003).
- Praitis, V., Casey, E., Collar, D. & Austin, J. Creation of low-copy integrated transgenic lines in *Caenorhabditis elegans*. *Genetics* **157**, 1217–1226 (2001).
- Mello, C. C., Kramer, J. M., Stinchcomb, D. T. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
- Strome, S. Fluorescence visualization of the distribution of microfilaments in gonads and early embryos of the nematode *Caenorhabditis elegans*. *J. Cell Biol.* **103**, 2241–2252 (1986).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank K. S. Ravichandran, M. Spector, T. Gumienny and L. Stergiou for comments on this manuscript, S. Milstein for the use of the *glu-1*(*op234*) allele, O. Hobert for the gift of the *P_{lim-7}* plasmid, J. Austin for pDPM0016 (*unc-119*(+)), and P. Gislser

for worm freezing. Some strains were supplied by the *Caenorhabditis* Genetic Center (CGC), which is supported by the National Institute of Health National Center for Research Resources. This work was supported by grants from the Swiss National Science Foundation, The Ernst Hadorn Foundation and the European Union (FP5 project APOCLEAR) to M.O.H., and by the Max Planck Society and an EU TMR Research Network grant to R.S.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.O.H. (Michael.Hengartner@molbio.unizh.ch).

Phospholipase C γ 1 controls surface expression of TRPC3 through an intermolecular PH domain

Damian B. van Rossum^{1*}, Randen L. Patterson^{4*}, Sumit Sharma¹, Roxanne K. Barrow¹, Michael Kornberg¹, Donald L. Gill⁵ & Solomon H. Snyder^{1,2,3}

¹Departments of Neuroscience, ²Pharmacology and Molecular Science, and

³Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

⁴Department of Biology, The Pennsylvania State University, State College, Pennsylvania 16802, USA

⁵Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, Maryland 21201, USA

* These authors contributed equally to this work

Many ion channels are regulated by lipids^{1–3}, but prominent motifs for lipid binding have not been identified in most ion channels. Recently, we reported that phospholipase C γ 1 (PLC- γ 1) binds to and regulates TRPC3 channels⁴, components

of agonist-induced Ca²⁺ entry into cells. This interaction requires a domain in PLC- γ 1 that includes a partial pleckstrin homology (PH) domain—a consensus lipid-binding and protein-binding sequence^{5,6}. We have developed a gestalt algorithm to detect hitherto ‘invisible’ PH and PH-like domains, and now report that the partial PH domain of PLC- γ 1 interacts with a complementary partial PH-like domain in TRPC3 to elicit lipid binding and cell-surface expression of TRPC3. Our findings imply a far greater abundance of PH domains than previously appreciated, and suggest that intermolecular PH-like domains represent a widespread signalling mode.

An amino-terminal portion of TRPC3 binds to PLC- γ 1 via a sequence of PLC- γ 1 that includes the SH3 domain and the carboxy-terminal half of a split PH domain (PH-c)⁴. To identify the interacting portion of PLC- γ 1 and TRPC3, we conducted a yeast two-hybrid analysis (Fig. 1a). Binding of PLC- γ 1 to TRPC3 requires amino acids 40–46 of TRPC3, with point mutations in this area abolishing binding. TRPC3 binds to PLC- γ 1 PH-c, not the SH3 domain. Moreover, TRPC3 binding seems to be specific for PLC- γ 1 PH-c, as the PH-c domain of AGAP1 does not bind TRPC3, and a single point mutation (F43A) in TRPC3 abolishes β -galactosidase activity in an alternative yeast two-hybrid system (Fig. 1b). *In vitro* protein binding experiments show that PLC- γ 1–TRPC3 interactions are dependent upon amino acids 40–46 of TRPC3 (Fig. 1c, d).

PH domain interactions typically involve full-length PH domains, so the binding of TRPC3 via the PH-c domain of PLC- γ 1 was perplexing. We wondered whether the PH-c domain of PLC- γ 1 might interact with a complementary PH-n domain in TRPC3. Available protein domain search programs (for example, BLAST CDD⁷, Pfam⁸, Prosite⁹ and SMART¹⁰) fail to detect PH domain consensus sequences in TRPC3. However, PH domains that are not recognized by conventional programs exist, exemplified by crystallographically defined PH domains in neurobeachin¹¹, TFIIH¹² and RanBP2 (ref. 13). To detect the PH domains in these proteins, we developed a gestalt algorithm to allow greater

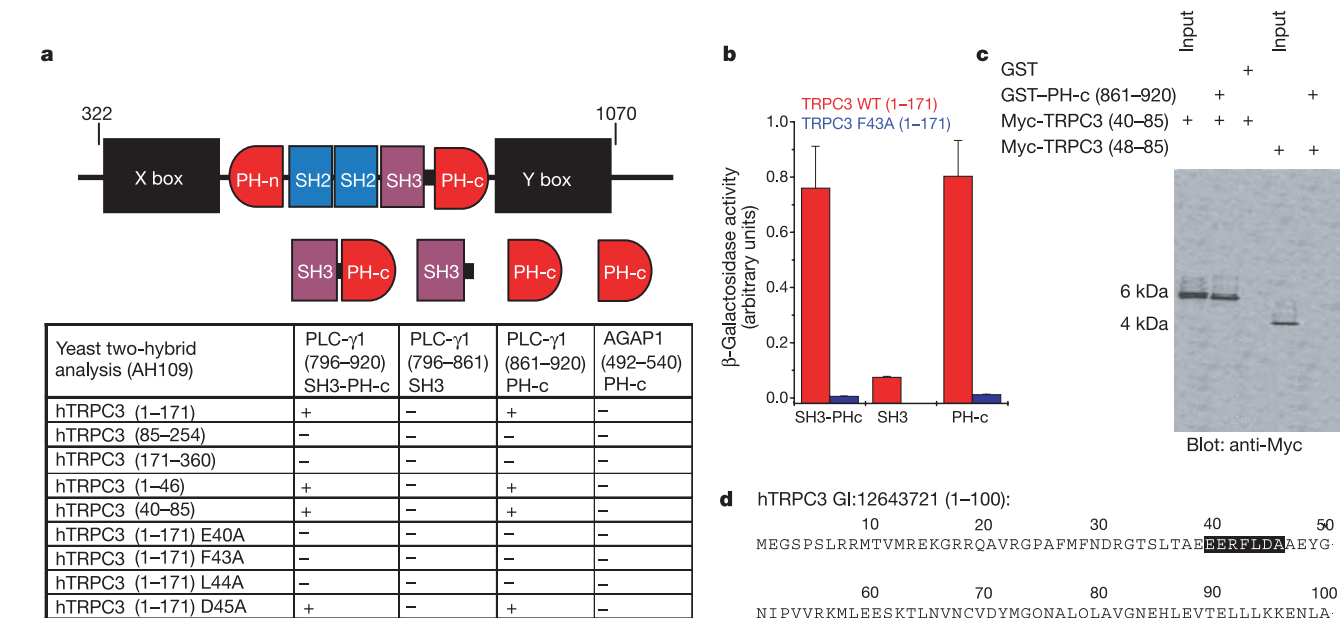


Figure 1 The PLC- γ 1 PH-c domain binds directly to amino acids 40–46 in TRPC3. **a**, Schematic depiction of the catalytic region of PLC- γ . The table depicts the results of yeast two-hybrid screening of rat PLC- γ 1 and human TRPC3 by X- α -Gal blue/white selection in AH109 yeast. **b**, β -Galactosidase activity (1 unit = 1 μ mol of ONPG to *o*-nitrophenol and *D*-galactose per min per cell) of pGBKT7-TRPC3 (wild type) or F43A co-transfected with pGADT7-PLC- γ 1 fragments in Y187 yeast. Error bars are standard

error of the mean. **c**, GST pull-down assay of 100 μ g of HEK293 cell lysates expressing Myc-tagged TRPC3 fragments 40–85 or 48–85, incubated with GST alone or GST-PLC- γ PH-c, run on SDS-PAGE and visualized by western blot analysis with anti-Myc. Input lanes are 10 μ g. **d**, hTRPC3 N-terminal amino acids 1–100 with the PLC- γ 1 binding site shaded.

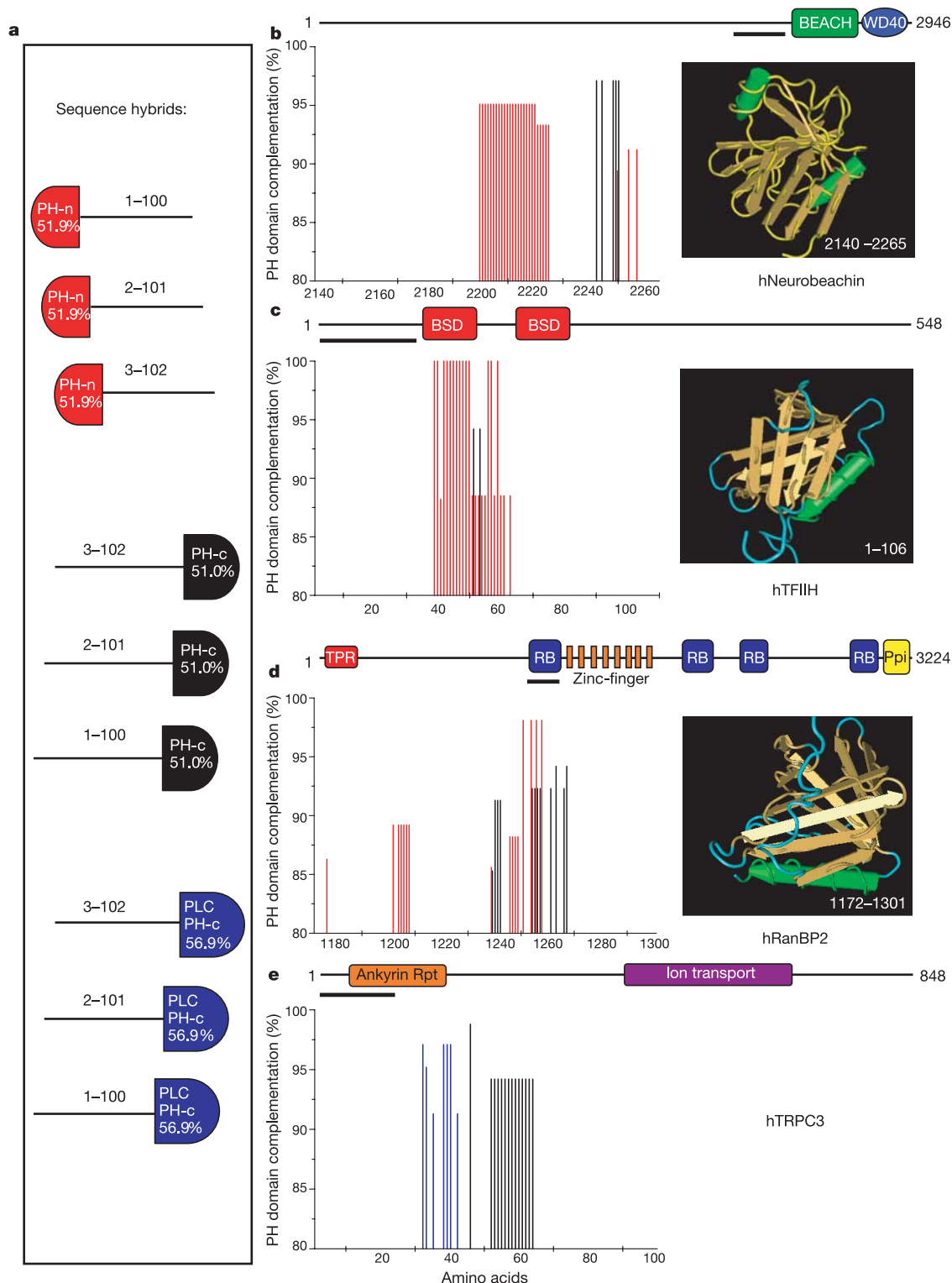


Figure 2 TRPC3 possesses an invisible PH-n domain at the site of PLC- γ 1 PH-c binding. **a**, Schematic depiction of the N-terminal PH consensus hybrids (red), C-terminal PH consensus hybrids (black), or C-terminal PLC- γ 1 PH-c sequence hybrids (blue). **b**, Blast-derived domain architecture of full-length neurobeachin with the region of interest underlined (amino acids 2140–2265; top panel). The left panel shows a graph of PH domain complementation in both N- (red) and C-terminal (black) hybrid directions for each amino acid position. The right panel shows the three-dimensional structure of the

corresponding area, from NCBI's Entrez structure database. **c**, Same as **b** but for hTFIIH (amino acids 1–106). BSD, domain in transcription factors and synapse-associated proteins. **d**, Same as **b** but for hRanBP2 (amino acids 1172–1301). TPR, tetratricopeptide repeats; RB, ran-binding domain; Ppi, peptidyl-prolyl *cis-trans* isomerase. **e**, Same as **b** but for hTRPC3 (amino acids 1–100), with the difference that PLC- γ 1 PH-c sequence hybrids are also plotted.

sequence divergence than conventional methods, which might then be applied to TRPC3. Briefly, by inserting either the N-terminal or C-terminal half of the PH domain consensus into each amino acid position of the target protein we essentially slide a partial PH consensus through the entire coding sequence (Fig. 2a, cartoon). At each sequence-hybrid position we assess PH domain complementation using NCBI Conserved Domain Search⁷ (see Methods). For neurobeachin, in the region containing the PH domain identified by X-ray crystallography (amino acids 2140–2265), we detected up to 95% complementation of a full PH domain whether scanning from the N-terminal (red) or C-terminal direction (black) (Fig. 2b). Similar results are obtained for TFIH and RanBP2 (amino acids 1–106 and 1172–1301, respectively) (Fig. 2c, d). Screening the entire amino acid sequence of neurobeachin, TFIH and RanBP2 we detected additional areas of complementation, which might represent other full or partial PH domains (data not shown).

Our studies with neurobeachin, TFIH and RanBP2 establish the capacity of the algorithm to identify PH domains not predicted by conventional methods. We used this algorithm to seek PH domain sequences in TRPC3 (Fig. 2e). Using consensus PH domain hybrids we detected prominent complementation when scanning from the C-terminal direction (black) (amino acids 50–60) but not from the N-terminal direction. Importantly, scanning TRPC3 with the PH-c sequence of PLC- γ 1 (blue) provides complementation around amino acids 40–48, corresponding exactly to the sequence that binds PLC- γ 1 (see Fig. 1d). For clarity, a raw NCBI Conserved Domain Search derived for a single TRPC3 sequence hybrid is provided (Supplementary Fig. 1). Thus, an intermolecular PH or PH-like domain may exist between PLC- γ 1 PH-c and TRPC3.

Support for TRPC3 possessing a PH-n without a PH-c domain comes from analyses on *smallwing* (*Drosophila* PLC- γ), alpha-1 syntrophin and syngap, where both halves of an 'invisible', or split intramolecular PH domain are identifiable (Supplementary Fig. 2).

To assess the functional relevance of the TRPC3–PLC- γ 1 intermolecular PH-like domain, we explored lipid binding, a property of many PH domains^{5,6} (Fig. 3). We generated *in vitro* translated, [³⁵S]methionine-labelled PLC- γ 1 PH-c (861–920), wild-type TRPC3 (1–171) or mutant (F43A) TRPC3 protein, confirmed expression by autoradiography (data not shown), and monitored lipid binding of these proteins alone or in combination on a 15-lipid-strip array. The combination of PLC- γ 1 PH-c and wild-type TRPC3 binds to a pattern of lipids that is distinct from binding observed with PLC- γ 1 PH-c or wild-type TRPC3 alone, most notably phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂; Fig. 3a, middle panel) and sphingosine-1-phosphate (S1P; data not shown). By contrast, lipid binding by the combination of TRPC3 F43A and PLC- γ 1 PH-c does not differ from PH-c alone, and binding to PtdIns(3,4)P₂ is undetected for all conditions (data not shown). As prototypical PH domains are not established to bind to S1P, binding by the putative TRPC3–PLC- γ 1 intermolecular domain could be due to either a distinct surface/pocket specific to PH-like domains or conformational changes in lipid structure from nitrocellulose immobilization.

To seek physiological relevance we explored TRPC3 Ca²⁺ channel activity *in vivo*. We monitored Fura-2AM-loaded HEK293 cells containing exogenously expressed full-length, Myc-tagged wild-type TRPC3 or the F43A mutant that does not bind PLC- γ 1. Overexpressed TRPC3 can be discriminated from endogenous

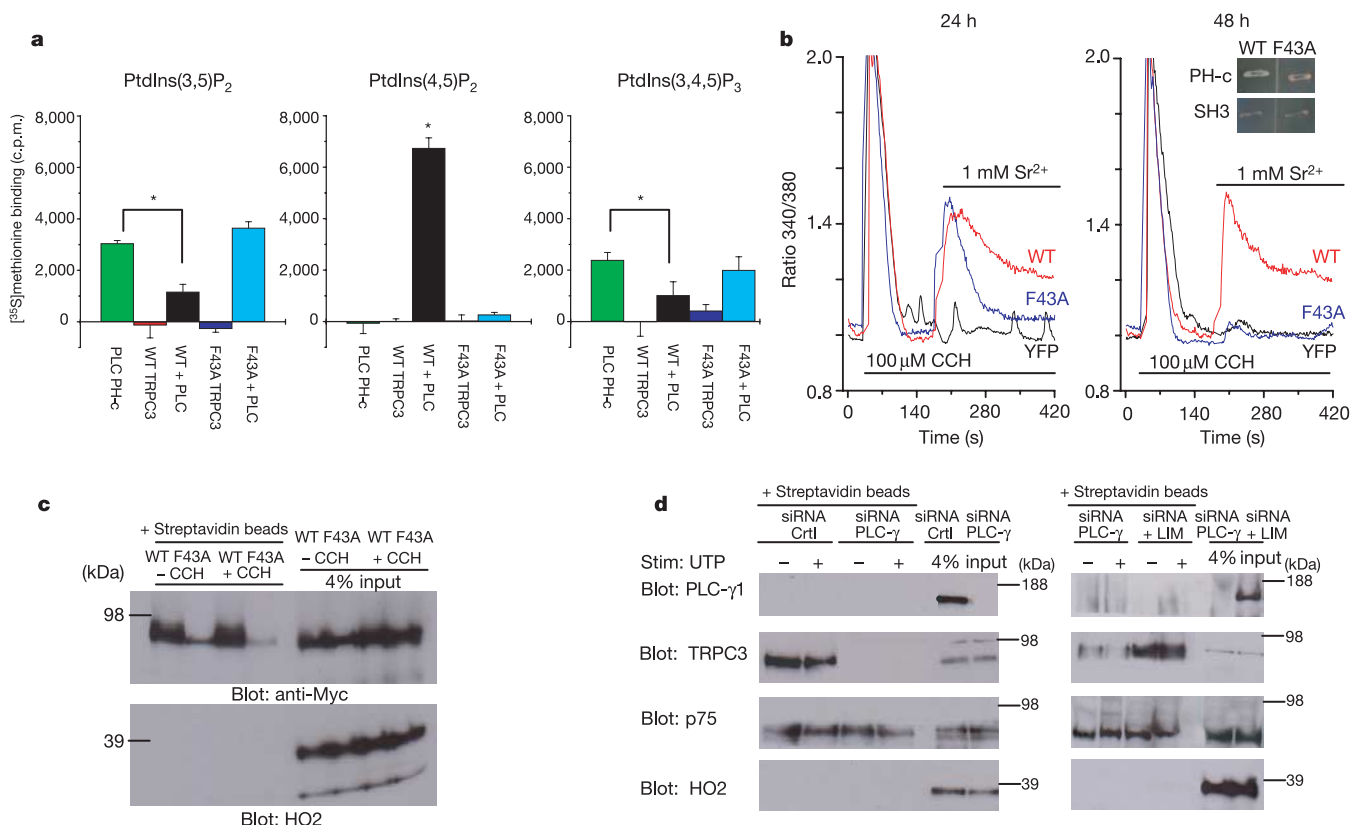


Figure 3 Interaction between PLC- γ 1 PH-c and TRPC3 confers lipid binding and affects membrane expression. **a**, [³⁵S]methionine counts from *in vitro* translated protein(s) incubated over PIP strips. c.p.m., counts per minute. Error bars are standard deviation and *P* value (*) ≤ 0.01 . **b**, Free Ca²⁺ measurements in HEK293 cells transfected for 24 or 48 h with YFP, YFP plus Myc-TRPC3 (wild type) or YFP plus F43A. Ca²⁺ pools were released by carbachol (CCH; first bar) in Ca²⁺-free medium followed by replacement with

Sr²⁺ (second bar). The inset shows yeast growth on selective media. **c**, Western blot of biotinylated HEK293 cells transfected with wild-type TRPC3 or the F43A mutant for 48 h, treated with or without CCH (100 μM). Input lanes 20 μg. **d**, Western blot of biotinylated rat PC12 cells transfected with scrambled siRNA or rPLC- γ 1 siRNA with or without hPLC- γ 1 lipase-inactive mutant (LIM) and UTP (100 μM). Input lanes 20 μg.

TRPC channels by replacing Ca^{2+} (1 mM) in the bathing medium with Sr^{2+} (1 mM), the flux of which selectively uses the over-expressed channels^{4,14}. At 24 h posttransfection, TRPC3 channel activity is the same for wild type and F43A (Fig. 3b, left). By contrast, at 48 h mutant channel activity is abolished, whereas robust wild-type channel activity persists (Fig. 3b, right).

As the F43A channel cannot bind PLC- γ 1, its channel gating is independent of PLC- γ 1. Therefore, we wondered how PLC- γ 1 was regulating TRPC3 activation through G_q -coupled muscarinic receptor stimulation. One possibility is protein stability; however, we detected little difference in expression between the wild-type and F43A TRPC3 channel at 48 h (Fig. 3c). The normal channel activity of the F43A mutant at 24 h indicates that it can localize to the plasma membrane. This finding, combined with the normal total level of mutant TRPC3 at 48 h, implies a defect in cell-surface expression of the channel at 48 h. Accordingly, we examined the cell-surface disposition of overexpressed wild-type and mutant channels at 48 h using cell-surface biotinylation (Fig. 3c). Whereas wild-type TRPC3 is expressed in the plasma membrane at 48 h, levels of the mutant channel are significantly reduced, with total protein levels being unaffected. We ascertained whether short interfering (si)RNA-induced depletion of PLC- γ 1 would influence cell-surface

levels of endogenous TRPC3 (Fig. 3d). Depletion of PLC- γ 1 reduces endogenous TRPC3 in the plasma membrane whereas a scrambled siRNA control does not. Regulation of Ca^{2+} entry by PLC- γ 1 is independent of its lipase function⁴. Rescued expression of human PLC- γ 1 lipase-inactive mutant restores plasma membrane expression of endogenous TRPC3 in the siRNA preparation (Fig. 3d, right).

These findings imply that TRPC3–PLC- γ 1 intermolecular PH-like domain formation regulates surface expression of TRPC3. However, these conclusions are based solely upon the use of the F43A mutant. From this alone one cannot conclude that formation of the predicted PH-like domain is responsible for the observed $\text{PtdIns}(4,5)\text{P}_2$ binding. Similarly, one cannot conclude that $\text{PtdIns}(4,5)\text{P}_2$ binding is important for function, as the F43A mutation disrupts the TRPC3–PLC- γ 1 complex. To investigate both questions requires the construction of a TRPC3 mutant that retains the ability to interact with PLC- γ 1 and generate the PH-like domain, but cannot bind $\text{PtdIns}(4,5)\text{P}_2$. TRPC3 contains eight conserved arginine or lysine residues in the predicted B1–B3 loop (amino acids 1–39, see Fig. 1d). Taking advantage of PLC- γ 1 PH-c complementation within wild-type TRPC3 at seven discrete points (blue lines in Fig. 2e), we discriminated which residues may be

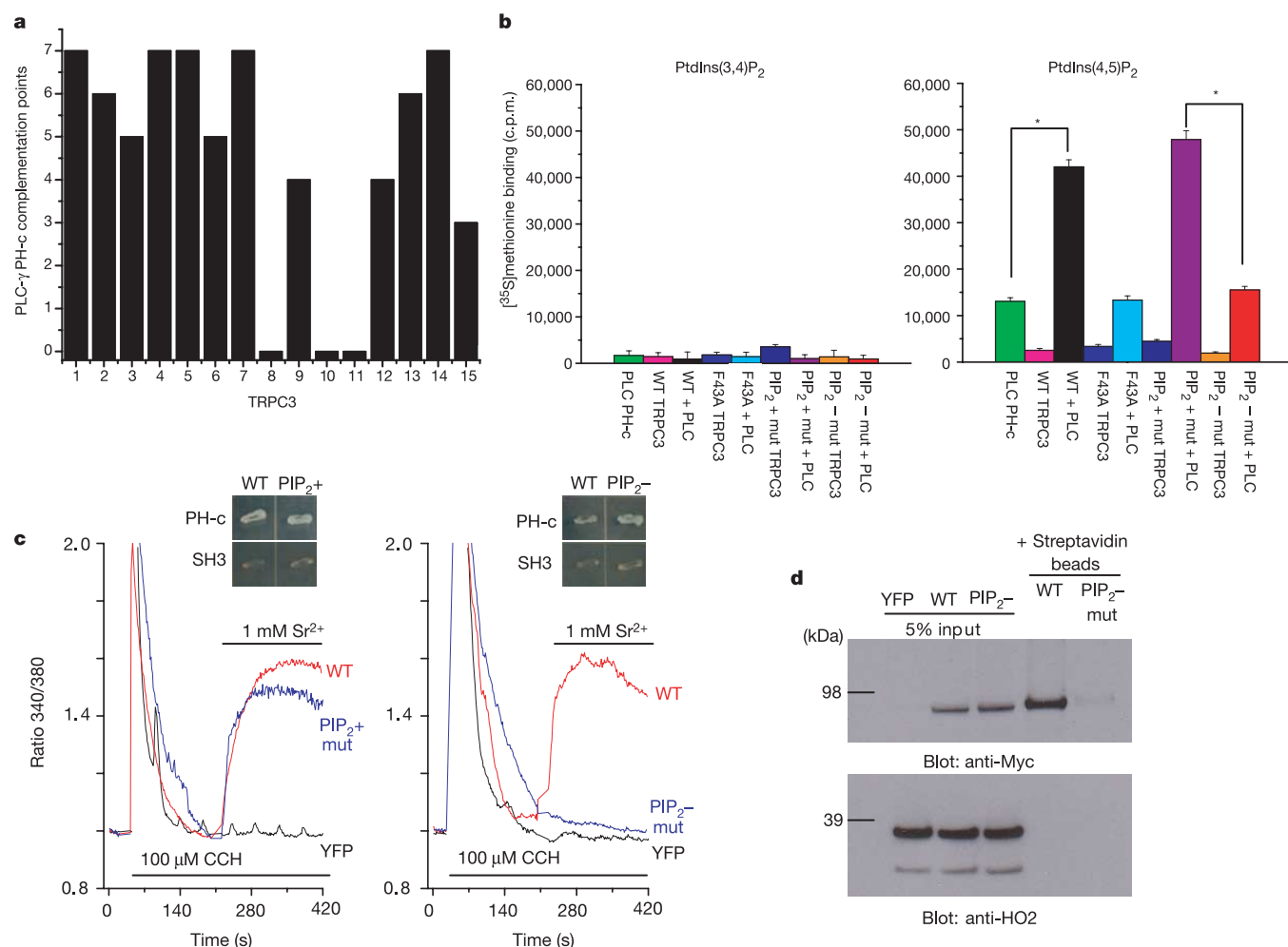


Figure 4 Intermolecular PH-like domain formation controls the surface expression of TRPC3. **a**, *In silico* scanning alanine mutagenesis predicts TRPC3 mutants retaining (PIP₂+mut) and lacking (PIP₂-mut) lipid binding. Lanes: 1, WT; 2, R8A; 3, R9A; 4, R8A/R9A (PIP₂+mut); 5, R14A; 6, K16A; 7, R18A; 8, R19A; 9, R14A/K16A; 10, R14A/K16A/R18A; 11, R14A/K16A/R18A/R19A (PIP₂-mut); 12, R23A; 13, R32A; 14, R42A; 15, R23A/R32A/R42A. **b**, $[^{35}\text{S}]$ methionine counts from *in vitro* translated protein(s) incubated

over lipid-conjugated agarose beads. Error bars are s.d. and P value (*) ≤ 0.01 . **c**, Free Ca^{2+} measurements in HEK293 cells transfected for 24 h with YFP alone or YFP plus Myc-TRPC3 (wild type), PIP₂+mut, or PIP₂-mut. Ca^{2+} pools were released in cells by CCH (first bar) followed by replacement with Sr^{2+} medium (second bar). The insets show yeast growth on selective media. **d**, Western blot of biotinylated HEK293 cells transfected with YFP alone, Myc-tagged TRPC3 (wild type) or PIP₂-mut for 24 h. Input lanes 20 μg .

involved in PtdIns(4,5)P₂ binding by conducting *in silico* scanning alanine mutagenesis (Fig. 4a). PH domain complementation by wild-type TRPC3 and R8A/R9A (PIP₂+mut) are similar; however, within the TRPC3 mutant R14A/K16A/R18A/R19A (PIP₂-mut) complementation is abolished (Fig. 4a). For both TRPC3 mutants (1–171), binding to PLC-γ1 PH-c (861–940) was confirmed by yeast two-hybrid analysis (Fig. 4c, inset).

We used PIP₂-conjugated agarose beads as an independent method to assay lipid binding. Specificity of binding was compared between PtdIns(3,4)P₂ and PtdIns(4,5)P₂, as these lipids have similar charge distribution. We generated *in vitro* translated, [³⁵S]methionine-labelled PLC-γ1 PH-c (861–940), wild-type TRPC3 (1–171), F43A, PIP₂+mut and PIP₂-mut, and confirmed expression by autoradiography (data not shown). We monitored lipid binding of these proteins alone or in combination. The combination of PLC-γ1 PH-c and wild-type TRPC3 or PIP₂+mut binds to PtdIns(4,5)P₂, with little to no PtdIns(3,4)P₂ binding (Fig. 4b). By contrast, the combination of PLC-γ1 PH-c and TRPC3 mutants F43A or PIP₂-mut does not differ compared with PH-c alone. This establishes a lipid-binding function for the TRPC3–PLC-γ1 intermolecular PH-like domain, and is the first direct evidence of lipid binding within the TRP superfamily.

To ascertain whether lipid binding modulates surface expression we monitored Fura-2AM-loaded HEK293 cells containing exogenously expressed full-length, Myc-tagged wild-type TRPC3, PIP₂+mut, or PIP₂-mut. At 24 h post-transfection, channel activity is the same for wild-type TRPC3 and PIP₂+mut (Fig. 4c, left). By contrast, at 24 h the activity of the TRPC3 PIP₂-mut channel is abolished (Fig. 4c, right). We examined the cell-surface disposition of overexpressed wild-type and PIP₂-mut channels at 24 h using cell-surface biotinylation (Fig. 4d). Whereas wild-type TRPC3 is expressed in the plasma membrane, PIP₂-mut channel levels are markedly reduced in the plasma membrane, with total protein levels being unaffected.

We wondered whether other instances of lipid binding might involve hitherto unrecognized PH-like domains. One study² identified a novel consensus sequence in TRPV1 and certain potassium channels that mediate PIP₂ modulation. Our algorithm for PH-like domains reveals complementation in the exact area (amino acids 780–819) known to mediate PIP₂ regulation of TRPV1. Furthermore, in yeast two-hybrid analysis, PLC-γ1 PH-c domain binds to rat (r)TRPV1 in a similar area to that of TRPC3 (peptide EEVQL versus EERFL, respectively), suggesting functional symmetry between these two channels (Supplementary Fig. 3).

Our findings provide a molecular mechanism for the regulation of agonist-induced Ca²⁺ entry by PLC-γ1 (ref. 4). Specifically, PLC-γ1 forms an intermolecular PH-like domain with TRPC3, regulating lipid binding and surface expression of TRPC3. Evidence includes: (1) an algorithm that identifies ‘invisible’ PH and PH-like domains; (2) elucidation of such a domain in TRPC3, which, with PLC-γ1 PH-c, binds to PtdIns(4,5)P₂ *in vitro*; (3) elimination of lipid binding by selective mutation of TRPC3 in either its PLC-γ1 PH-c binding site or its predicted B1–B3 loop; (4) abolition of TRPC3 activity by these selective mutations; and (5) blockade of cell-surface expression of endogenous TRPC3 by siRNA-induced depletion of PLC-γ1.

Our gestalt algorithm for identifying invisible PH domains can reveal previously unrecognized PH domains, exemplified by its detection of PH domains in neurobeachin, TFIH and RanBP2, which could only be realized through X-ray crystallography. Furthermore, this algorithm, a partial consensus sequence that slides across a protein's amino acid sequence, may identify other hidden domains as well. □

Methods

Yeast two-hybrid analysis

Experiments were performed with the Matchmaker 3 yeast two-hybrid system (Clontech),

following the manufacturer's instructions, with pGBKT7-BD-TRPC3 fragments and pGADT7-AD-PLC-γ1 or AGAP fragments (where AD and BD indicate β-galactosidase-activating domain and -binding domain, respectively). Human (h)TRPC3 mutations were introduced using the QuikChange site-directed mutagenesis kit (Stratagene) and confirmed by sequencing.

GST pull down

Cell lysis buffer (150 mM NaCl, 50 mM Tris, pH 7.8, 1% Triton X-100, 1 mM EDTA) was added to 100 μg of Myc-TRPC3 40–85- or 46–85-expressing HEK293 whole-cell lysates (total volume 500 μl). GST–Sephacrose beads and purified GST–PLC-γ1 PH-c or GST were added, incubated on a rotator for 1 h at 4 °C, washed three times with lysis buffer and quenched with 20 μl of SDS sample buffer. Co-precipitates were resolved by SDS–PAGE and analysed by western blot analysis (input lanes 10 μg).

Domain alignments

We generated serial hybrid sequences comprised of a target sequence (for example, human neurobeachin GI:21536252, hTFIH GI:416727, hRanBP2 GI:6382079 or hTRPC3 GI:12643721) and a partial PH domain ‘donor’ sequence inserted at each target amino acid position either N-terminally or C-terminally. PH-n front consensus, VIKEGWLLKKSSGGKSSWKRYFVLFNGV LLYYKSKKKSSSKPKGSIPLSGCT; PH-c back consensus, GCTVREAPDSDSDKKKNCFEIVTPDRKTLTLQAEESEEEVVEALR KAIKAL; and rPLC-γ1 (GI:6981370) PH-c sequence (865–930), NSPLGDLRLGVLDV PACQAIIRPEGGKNNRLVFV SISMPSSVAQWSDVAADSQEELQDWVKKIREVA.

Domain searches were performed using NCBI Conserved Domain Database (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) on overlapping, 150-amino-acid hybrid fragments (settings: search database CDDv2.00-11382 PSSMS, expect 0.01, no filter, search mode multiple hits 1-pass). Significant PH domain complementation was monitored as a change in the SMART or PFAM conserved domain length greater than or equal to 85% (because scores for neurobeachin, TFIH and RanBP2 did not fall below this cutoff) and was plotted on the y axis at the respective target amino acid hybrid point.

In vitro translation

In vitro translation of BD-TRPC3 and AD-PLC-γ vectors was performed using Promegas TNT quick-coupled transcription/translation systems following the manufacturer's instructions.

Lipid binding

PIP strips (Echelon P-6001) cut into individual spots were blocked in Tris-buffered saline with 0.1% Tween and 3% fatty-acid-free bovine serum albumin (Sigma) (TBST-3% BSA) for 1 h. Each spot was incubated for 3 h at room temperature with 500 μl of TBST-3% BSA plus 10 μM of the *in vitro* translated protein(s) of interest, washed six times in 1 ml of TBST-3% BSA, dissolved in 6 ml of scintillation fluid overnight and counted. The blank spot was used to subtract background from the remainder of the samples. A total of 20 μl bed volume of PIP₂ beads was equilibrated in 1 ml of TBS with 0.5% Tween overnight. Equilibration buffer was removed and replaced with 200 μl of 0.5% TBST containing 10 μM of the appropriate [³⁵S]methionine-labelled construct(s) and incubated for 2 h at room temperature, washed ten times in 1 ml 0.5% TBST, transferred to scintillation fluid and counted. Background counts obtained from PIP₂ beads incubated with mock-translated rabbit reticulocyte lysate containing empty vector and [³⁵S]methionine were subtracted from all samples. In control experiments (data not shown), [³⁵S]methionine-labelled β-galactosidase-activating domain (AD), -binding domain (BD), or AD plus BD failed to bind to any lipid spot or bead. Each experiment was performed a minimum of three times, and the data expressed in the bar graphs are averages of triplicates obtained from one experiment. Error bars represent the standard deviation from one experiment and all significance is ≤0.01.

Calcium measurements

Ca²⁺ measurements were as described^{4,14}. All traces are averages from multiple (30–50) cells and are representative of at least three individual experiments. Fluorescence emission at 505 nm was monitored with excitation at 340 and 380 nm; Ca²⁺ measurements are shown as 340/380-nm ratios obtained from averages of single cells.

TRPC3 antibody

Rabbit polyclonal antiserum against TRPC3 was generated by injecting New Zealand white male rabbits with the peptide MREKGRQAVRGPAFMFNDRC coupled to keyhole limpet haemocyanin (Pierce Biotechnology). Initial injection was with complete Freund's adjuvant, and boost injections at days 14, 21 and 49 were with incomplete Freund's adjuvant. Rabbit injections, bleeds and housing were performed by Cocalico Biologicals. Antibodies were affinity purified using the antibody peptide coupled to the Pierce Biotechnology Ultralink iodoacetyl column, as per the manufacturer's instructions.

Cell-surface biotinylation

HEK293 cells were transfected with Myc-TRPC3 constructs for 24 or 48 h, and PC12 cells with either scrambled, PLC-γ1 siRNA⁴, or siRNA plus Flag-tagged, lipase-inactive human PLC-γ1 for 48 h. Cells were washed three times in 10 ml ice-cold PBS pH 8, and incubated in 6 ml of 1 mM Sulpho-NHS-LC-biotin (Pierce) in PBS pH 8 for 3 h. Cells were quenched with 10 ml of PBS pH 8, 10 mM Tris and 100 mM glycine for 10 min, washed twice in 10 ml PBS pH 8, lysed (500 μl RIPA buffer) and the protein assay performed. A total of 500 μg of cell lysate was dissolved in 500 μl of RIPA buffer plus 25 μl bed volume of streptavidin-agarose beads, and rocked overnight. Beads were washed ten times in 1 ml RIPA buffer, boiled in 30 μl sample buffer, and analysed by SDS–PAGE and western blot analysis. Polyclonal anti-HO2 antibody¹⁵ served as a negative control for surface labelling, and

p75 (a surface receptor) served as a cell-surface control for siRNA-induced PLC- γ 1 depletion.

Reagents

Enhanced yellow fluorescent protein (YFP) vector and Lipofectamine were from Clontech; anti-Myc, [35 S]methionine, carbachol, ONPG and GST-Sepharose were from Sigma; Fura-2/acetoxymethyl ester was from Molecular Probes. siRNA duplexes were from Dharmacon Research. Anti-p75 antibody was from Upstate.

Received 18 August; accepted 17 December 2004; doi:10.1038/nature03340.

1. Runnels, L. W., Yue, L. & Clapham, D. E. The TRPM7 channel is inactivated by PIP(2) hydrolysis. *Nature Cell Biol.* **4**, 329–336 (2002).
2. Prescott, E. D. & Julius, D. A modular PIP2 binding site as a determinant of capsaicin receptor sensitivity. *Science* **300**, 1284–1288 (2003).
3. Suh, B. C. & Hille, B. Recovery from muscarinic modulation of M current channels requires phosphatidylinositol 4,5-bisphosphate synthesis. *Neuron* **35**, 507–520 (2002).
4. Patterson, R. L. *et al.* Phospholipase C- γ is required for agonist-induced Ca^{2+} entry. *Cell* **111**, 529–541 (2002).
5. DiNitto, J. P., Cronin, T. C. & Lambright, D. G. Membrane recognition and targeting by lipid-binding domains. *Sci. STKE* **re16** (2003).
6. Rebecchi, M. J. & Scarlata, S. Pleckstrin homology domains: a common fold with diverse functions. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 503–528 (1998).
7. Marchler-Bauer, A. *et al.* CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* **31**, 383–387 (2003).
8. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
9. Gattiker, A., Gasteiger, E. & Bairoch, A. ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl. Bioinform.* **1**, 107–108 (2002).
10. Letunic, I. *et al.* SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* **32**, D142–D144 (2004).
11. Jögl, G. *et al.* Crystal structure of the BEACH domain reveals an unusual fold and extensive association with a novel PH domain. *EMBO J.* **21**, 4785–4795 (2002).
12. Gervais, V. *et al.* TFIIH contains a PH domain involved in DNA nucleotide excision repair. *Nature Struct. Mol. Biol.* **11**, 616–622 (2004).
13. Vetter, I. R., Nowak, C., Nishimoto, T., Kuhlmann, J. & Wittinghofer, A. Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature* **398**, 39–46 (1999).
14. Ma, H. T. *et al.* Requirement of the inositol trisphosphate receptor for activation of store-operated Ca^{2+} channels. *Science* **287**, 1647–1651 (2000).
15. Baranano, D. E. *et al.* A mammalian iron ATPase induced by iron. *J. Biol. Chem.* **275**, 15166–15173 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Boehning, G. Caraveo, J. Kendall, A. Resnick and R. E. Rothe for discussion; P.-G. Suh for the gift of the PLC- γ 1 antibody; and B. VanRossum for graphics. This research was supported by US Public Health Service Grants and a Research Scientist Award (to S.H.S.), a National Institutes of Health Grant (to D.L.G.), and National Research Service Awards (to R.L.P.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.H.S. (ssnyder@jhmi.edu).

Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates

Mart Loog & David O. Morgan

Department of Physiology, University of California, San Francisco, California 94143-2200, USA

Cell-cycle events are controlled by cyclin-dependent kinases (CDKs), whose periodic activation is driven by cyclins. Different cyclins promote distinct cell-cycle events, but the molecular basis for these differences remains unclear^{1,2}. Here we compare the specificity of two budding yeast cyclins, the S-phase cyclin Clb5 and the M-phase cyclin Clb2, in the phosphorylation of 150 Cdk1 (Cdc28) substrates. About 24% of these proteins were phosphorylated more efficiently by Clb5–Cdk1 than Clb2–Cdk1.

The Clb5-specific targets include several proteins (Sld2, Cdc6, Orc6, Mcm3 and Cdh1) involved in early S-phase events. Clb5 specificity depended on an interaction between a hydrophobic patch in Clb5 and a short sequence in the substrate (the RXL or Cy motif). Phosphorylation of Clb5-specific targets during S phase was reduced by replacing Clb5 with Clb2 or by mutating the substrate RXL motif, confirming the importance of Clb5 specificity *in vivo*. Although we did not identify any highly Clb2-specific substrates, we found that Clb2–Cdk1 possessed higher intrinsic kinase activity than Clb5–Cdk1, enabling efficient phosphorylation of a broad range of mitotic Cdk1 targets. Thus, Clb5 and Clb2 use distinct mechanisms to enhance the phosphorylation of S-phase and M-phase substrates.

A long-standing question in cell-cycle control is how different cyclins drive the distinct events of S phase and M phase^{1,2}. One model, termed the quantitative model of cyclin function, suggests that S phase is triggered by low levels of cyclin–CDK activity and M phase is initiated at higher levels of activity^{3,4}. According to this model, apparent differences in cyclin function are due primarily to differences in their timing and levels of expression. In contrast, the qualitative model proposes that different cyclins possess different intrinsic functional capacities, perhaps because they modulate the substrate specificity of the associated CDK or alter its subcellular location^{1,5}. Studies in budding yeast, for example, argue that the S-phase cyclin Clb5 possesses higher intrinsic S-phase-promoting activity than the M-phase cyclin Clb2 (refs 6, 7). Biochemical studies with mammalian cyclins have demonstrated cyclin specificity in the phosphorylation of a small number of substrates: for example, mammalian cyclin-A–CDK, but not cyclin-B–CDK, phosphorylates the pRb-related protein p107 (ref. 8). However, few cyclin-specific substrates have been identified in any species, and the general importance of cyclins in CDK substrate specificity remains unclear.

To assess the global importance of cyclin specificity in cell-cycle control, we measured the phosphorylation of a large number of Cdk1 substrates by S-phase and M-phase cyclin–CDK complexes. We recently identified 181 budding yeast Cdk1 substrates⁹, and in the present study we measured the kinase activities of Clb5–Cdk1 and Clb2–Cdk1 towards 150 of these proteins (obtained from proteomic libraries; see Methods). To prevent background phosphorylation by contaminating protein kinases in the reactions, we used a mutant form of Cdk1, Cdk1-as1, that contains an enlarged ATP-binding site. Purified Clb5–Cdk1-as1 or Clb2–Cdk1-as1 complexes were incubated with the test substrate and the bulky ATP analogue [γ - 32 P]N⁶-(benzyl)ATP, which only the mutant Cdk1-as1 enzyme can use⁹. Reactions were performed with amounts of Clb5–Cdk1-as1 and Clb2–Cdk1-as1 that possessed equal activities toward the non-specific substrate histone H1 (Fig. 1a). Phosphate incorporation was then divided by the amount of substrate protein, and the logarithm of this ratio was designated as the substrate P-score, as described previously⁹. All reactions were performed at very low substrate concentrations (presumably well below K_m); thus, differences in P-scores between the two kinases provide a reasonable estimate of relative k_{cat}/K_m values, where K_{cat} is catalytic constant and K_m is Michaelis constant.

All substrate P-scores for the two kinases are plotted in Fig. 1b. These data suggest that, on a histone H1-normalized scale, most of the substrates – about 110 of the 150 – are equally good substrates for Clb5–Cdk1 and Clb2–Cdk1, because they exhibit 2.5-fold or less specificity for either kinase and fall in the middle diagonal region of the plot. Most of the remaining substrates, falling to the right of the diagonal, are specific for Clb5. Among these were 14 substrates with specificity for Clb5 ranging from 10-fold to 800-fold, whereas 22 proteins displayed specificity of between 2.5-fold and 10-fold (note that the scale on this plot is logarithmic). The top Clb5-specific substrates included several proteins involved in DNA replication (Orc6, Orc2, Mcm3, Cdc6 and Sld2), spindle pole body function

(Mps2 and Spc110), APC activation (Cdh1) and other functions (Supplementary Table 1). Surprisingly, we found no highly Clb2-specific substrates and only a few substrates that displayed minimal (2.5–3.0-fold) specificity for Clb2.

We next determined the mechanism underlying Clb5 specificity. Substrate recognition by cyclin–CDK complexes is known to be governed primarily by an interaction between the substrate's consensus phosphorylation sequence, S/T*PXX/R, and the CDK active site¹⁰. In addition, a region on the surface of some cyclins, called the hydrophobic patch, has been reported to interact with a sequence motif called an RXL or Cy motif on some CDK substrates and inhibitors^{10–14}. Mutation of the hydrophobic patch decreases the biological activity, but not the histone H1 kinase activity, of the Clb5–Cdk1 complex¹⁵, but the broad significance of this interaction for CDK substrate targeting remains unclear. We therefore analysed the effects of hydrophobic patch mutations on the phosphorylation of Clb5-specific substrates identified in our screen. We found that the high Clb5 specificity of these substrates was entirely dependent on a functional hydrophobic patch (Fig. 2a). Mutation of the hydrophobic patch in Clb2 had no significant effect on its activity towards these substrates, and patch mutations in either cyclin had no effect on the phosphorylation of non-specific substrates such as histone H1 (Fig. 2a).

More detailed kinetic analyses were performed with two Clb5-specific substrates, the replication protein Cdc6 and the putative spindle protein Fin1 (ref. 16), which were chosen because they could be prepared in the large amounts needed for these studies. In both cases, the hydrophobic patch-dependent substrate interaction resulted in highly efficient phosphorylation, with K_m values in the low micromolar range (Fig. 2b).

We also attempted to identify substrate regions that interact with the hydrophobic patch of Clb5. With the use of site-directed mutagenesis, we screened potential RXL, RXF or KXL motifs in the Clb5-specific substrate Fin1 and found that a single KXL motif (residues 191–193) was almost entirely responsible for the high Clb5 specificity (Fig. 2c). Mutation of this sequence resulted in a 30-fold

decrease in k_{cat}/K_m with Clb5–Cdk1 (data not shown). The rate of Fin1 phosphorylation by Clb2–Cdk1 was not significantly affected by mutation of this motif (Fig. 2c) or any of the other related motifs in Fin1 (data not shown).

During the course of these studies, we noticed that purified Clb5–Cdk1 had a lower histone H1 kinase activity than Clb2–Cdk1. This difference was not due to differences in binding of the CDK inhibitor Sic1, because the kinase complexes were purified from a strain lacking Sic1. To address this issue further, we performed kinetic analyses with a peptide substrate that is derived from histone H1 and is ideal for characterization of the CDK active site without interference from other interactions. Purified Clb2–Cdk1 was about 10–20-fold more active towards this peptide or histone H1 than purified Clb5–Cdk1 (Table 1). The same was true for the mutant Cdk1-as1 complexes used for the proteomic screen described above (data not shown). The difference in activities was due almost entirely to a tenfold difference in K_m values for the peptide substrate (Table 1), whereas k_{cat} and K_m values for ATP (data not shown) were similar for both enzymes. A similar difference in peptide K_m was observed with Clb5- and Clb2-associated kinases immunoprecipitated from cell lysates (data not shown). These results, combined with the similarity in k_{cat} values for the two complexes, argue that the difference in activities of the two kinases was not due to partial inactivation of Clb5–Cdk1 during purification.

These data reveal a previously unrecorded principle of cyclin function: rather than simply activating a CDK, different cyclins can differentially modulate the intrinsic properties of the CDK active site. These data also suggest that, with the exception of specific substrates that interact with the Clb5 hydrophobic patch, Clb5–Cdk1 has lower activity than Clb2–Cdk1 towards the general substrates that lie along the diagonal of Fig. 1b. Thus, if the data in Fig. 1b had been obtained with equal kinase protein amounts rather than equal histone H1 kinase activities, P-scores for Clb2 would have increased by more than 1 unit, resulting in a majority of proteins that were above the diagonal and could therefore be considered Clb2-specific.

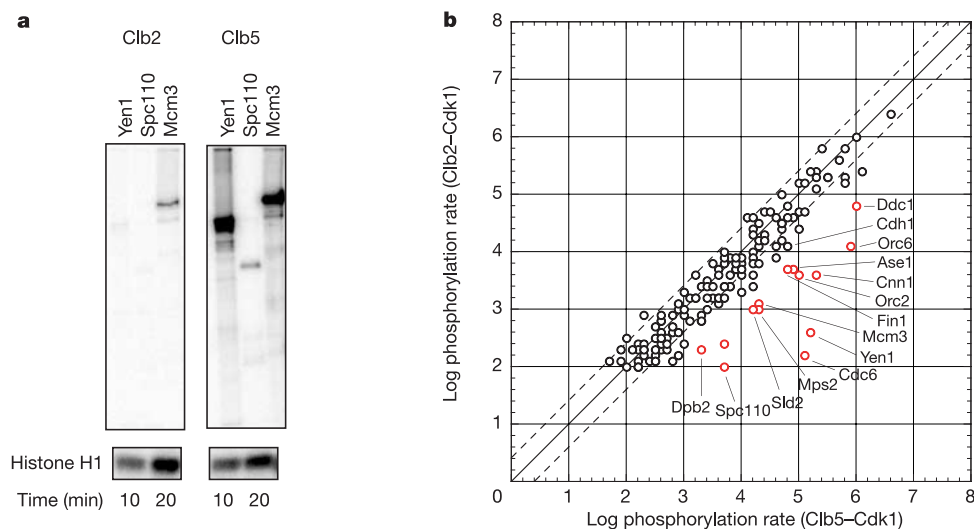


Figure 1 Identification of Clb5-specific Cdk1 substrates in budding yeast. **a**, Representative data showing phosphorylation of three Clb5-specific substrates. Purified Clb2–Cdk1-as1 (left) and Clb5–Cdk1-as1 (right) were normalized for histone H1 kinase activity (lower panels) and incubated with the indicated GST-tagged substrate proteins and [γ -³²P]N⁶-(benzyl)ATP. Reaction products were analysed by SDS–PAGE and autoradiography. **b**, Rates of phosphorylation of 150 substrates by Clb2–Cdk1 and Clb5–Cdk1 were divided by substrate amount and are plotted here as logarithmic values.

The activities of the two kinases were normalized to histone H1 kinase activity. The central diagonal region (bounded by dashed lines) contains substrates whose phosphorylation rates with the two kinases were similar (less than a 2.5-fold difference in phosphorylation rate). Red circles indicate 14 substrates that are phosphorylated more than tenfold more rapidly by Clb5–Cdk1 than by Clb2–Cdk1. Phosphorylation rates for all substrates are given in Supplementary Table 1.

Finally, to investigate the effects of cyclin specificity *in vivo*, we analysed the phosphorylation state of several Cdk1 substrates during the cell cycle in yeast strains in which the open reading frame of *CLB5* was replaced with that of *CLB2* at the *CLB5* locus. As shown previously⁶, the *clb5::CLB2* strain provides an excellent system in which to compare the intrinsic specificity of the two cyclins in the absence of normal differences in the timing of cyclin expression. Clb2 expression in the *clb5::CLB2* strain increased at the same time in the cell cycle as Clb5 in a wild-type strain (Fig. 3a). Clb2-associated histone H1 kinase activity in immunoprecipitates from the *clb5::CLB2* strain was about 3–4-fold greater than that of Clb5 from a wild-type strain (Fig. 3b). This difference is less than that seen with purified kinases, presumably due to relatively low Clb2 expression from the *CLB5* locus (Fig. 3a).

We first analysed the phosphorylation of the Clb5-specific substrate Fin1 after release from G1 arrest. Fin1 displays a Cdk1-dependent mobility shift in western blots⁹. Here we found that Fin1 in wild-type cells was fully phosphorylated immediately after its synthesis in early S phase, and was then dephosphorylated after degradation of Clb5 in mitosis (despite the continued presence of Clb2; Fig. 3c). However, in the *clb5::CLB2* strain, Fin1 phosphorylation during S phase was decreased by 30–40%. Similar results were obtained when Fin1 was replaced by a form in which the KXL motif was mutated. A decrease in phosphorylation of this magnitude is likely to reflect a large decrease in the rate of phosphorylation *in vivo*, as argued by theoretical considerations of the effects of changing kinase and phosphatase activities on substrate phosphorylation state¹⁷. For example, if the kinase and phosphatase acting on a substrate are not saturated with substrate (that is, under first-order conditions), then a 99-fold decrease in kinase activity is required for a decrease in substrate phosphorylation from 99% to 50%. Our data therefore indicate that Fin1 is a highly Clb5-specific substrate *in vivo* and that the interaction between the KXL motif of Fin1 and the

Table 1 Kinase activities of Clb2-Cdk1 and Clb5-Cdk1

Kinase	Substrate	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu\text{M}^{-1}\text{min}^{-1}$)
Clb2-Cdk1	Histone peptide	45.9 $\pm$ 15.4	189 $\pm$ 14	4.1
	Histone H1	—	—	4.9
	Fin1	—	—	6.4
Clb5-Cdk1	Histone peptide	521 $\pm$ 98	114 $\pm$ 9	0.22
	Histone H1	—	—	0.63
	Fin1	3.1 $\pm$ 0.5	102 $\pm$ 16	33.0

Kinetic properties of purified Clb2-Cdk1 and Clb5-Cdk1 with a histone H1-derived substrate peptide (PKTPKKAKKL) were compared with activity towards histone H1 and the Clb5-specific substrate Fin1. Virtually identical results were obtained with another commonly used peptide substrate (ADAQHATPPKKKRKVEDPKDF; data not shown). Values are means $\pm$ s.d.

hydrophobic patch of Clb5 is important for this specificity.

Similar results were obtained with the Clb5-specific substrate Sld2 (Fig. 3d). Replacement of Clb5 by Clb2 delayed the onset and decreased the amount of Sld2 phosphorylation, helping to explain previous evidence that the intrinsic S-phase-promoting activity of Clb2 in this strain is less than that of Clb5 (ref. 6). We also analysed a protein, Slk19, that exhibits no apparent specificity for Clb5 or Clb2 *in vitro*. Patterns of Slk19 phosphorylation were similar in wild-type and *clb5::CLB2* strains after release from G1 arrest (Fig. 3e), further indicating that the cyclin specificities observed *in vitro* (Fig. 1b) are relevant *in vivo*.

Thus, our large-scale comparative analysis of Cdk1 specificity unveiled a large group of highly specific protein targets for the S-phase cyclin Clb5. At the onset of S phase, this specificity might contribute to the switch-like phosphorylation of substrates required for efficient S-phase progression. Phosphorylation of Sld2, for example, is needed for the initiation of replication¹⁸, Cdh1 phosphorylation allows the accumulation of cyclins and other regulators¹⁹, and proteins of the pre-replicative complex (Cdc6, Orc2,

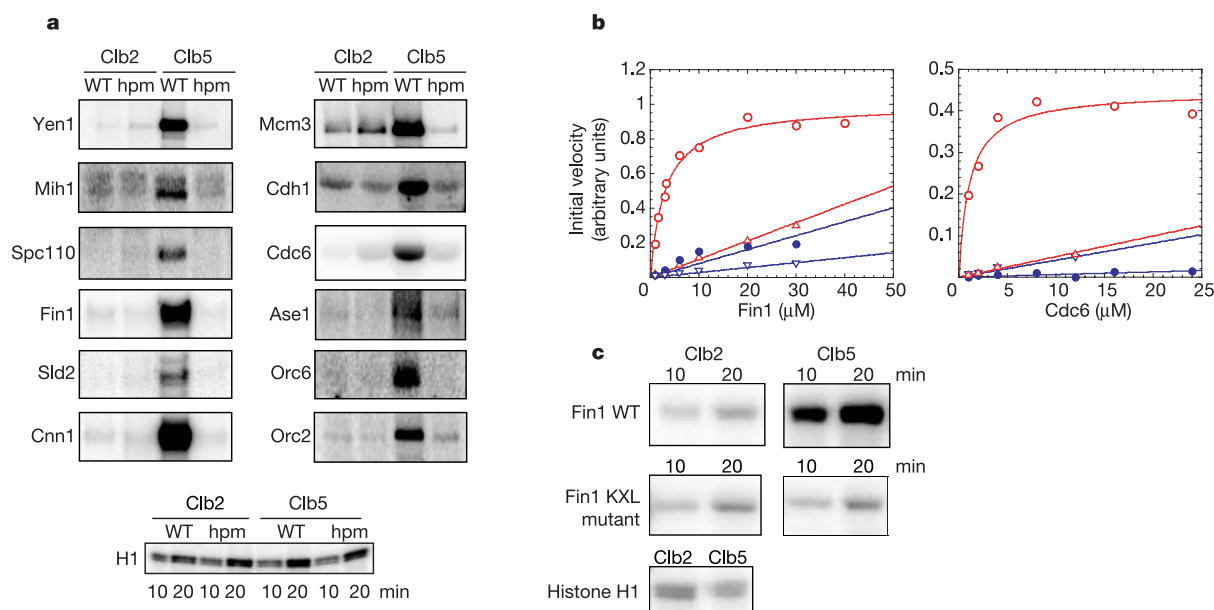


Figure 2 Clb5 specificity depends on an interaction between the Clb5 hydrophobic patch and an RXL motif in the substrate. **a**, Effect of the hydrophobic patch mutation (hpm) on the rate of Clb2-Cdk1- and Clb5-Cdk1-catalysed phosphorylation of substrates. Selected substrates with a high degree of Clb5 specificity were phosphorylated by purified Clb2-Cdk1, Clb2hpm-Cdk1, Clb5-Cdk1 and Clb5hpm-Cdk1. As a control, histone H1 phosphorylation is also shown at two time points (bottom panel). WT, wild type. **b**, Steady-state kinetic characterization of Clb5-specific substrates Fin1 (left) and Cdc6 (right), using

purified Clb2-Cdk1 (blue circles), Clb2hpm-Cdk1 (blue triangles), Clb5-Cdk1 (red circles) and Clb5hpm-Cdk1 (red triangles) (normalized for histone H1 kinase activity). For Fin1, $K_{m,\text{Clb5}} = 3.1 \mu\text{M}$; for Cdc6, $K_{m,\text{Clb5}} = 1.1 \mu\text{M}$. **c**, A KXL motif in Fin1 is responsible for the Clb5 specificity of Fin1 phosphorylation. A standard kinase assay, normalized for histone H1 kinase activity, was performed with wild-type Fin1 and with a KXL mutant in which residues Lys 191 and Leu 193 were changed to alanine.

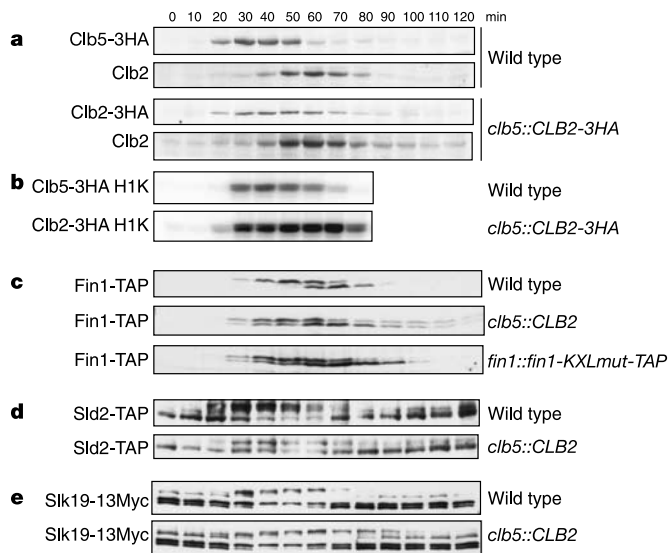


Figure 3 Clb5-specific substrate phosphorylation occurs *in vivo*. The indicated yeast strains were arrested in G1 with $1 \mu\text{g ml}^{-1}$ α -factor and then released from the arrest. New α -factor was added after the initiation of budding, to prevent entry into the next cell cycle. Cells were harvested for western blotting or kinase assays at the indicated times. **a**, western blotting of the indicated cyclins in lysates from a wild-type strain carrying *CLB5-3HA* at the *CLB5* locus, or a strain carrying *CLB2-3HA* at the *CLB5* locus (*clb5::CLB2-3HA*). **b**, Histone H1 kinase activity was measured in immunoprecipitates of Clb5-3HA from the wild-type strain and Clb2-3HA from the *clb5::CLB2-3HA* strain. **c**, **d**, Phosphorylation-dependent mobility shifts were measured by western blotting of the Clb5-specific substrates Fin1 (**c**) and Sld2 (**d**) (each TAP-tagged) in wild-type and *clb5::CLB2* strains. In both cases the upper band represents the phosphorylated form. A similar experiment was performed in a strain carrying the *fin1-KXL* mutation at the *FIN1* locus. Quantification of the two Fin1 bands at the 30–50-min time points revealed that Fin1 is 100% phosphorylated in wild-type cells, 63–68% phosphorylated in *clb5::CLB2* cells and 59–63% phosphorylated in *fin1-KXLmut* cells. **e**, Similar studies were performed with 13Myc-tagged Slk19, an example of a protein with no significant cyclin specificity. Slk19 migrates as two major bands, the lower of which is a proteolytic fragment produced by separase. Each of the two bands displays decreased mobility after phosphorylation⁹.

Orc6 and Mcm3) must be phosphorylated to prevent DNA re-replication²⁰. Consistent with this is the recent observation¹⁴ that Clb5 binding to an RXL motif in Orc6 is important for the suppression of DNA re-replication. Clb5 specificity might also be important in late mitosis. Clb5 is degraded in anaphase, before most Clb2 (Fig. 3a)^{21,22}, and it can be imagined that early dephosphorylation of Clb5-specific substrates might contribute in some way to the control of late mitotic events.

Our work also revealed that Clb2–Cdk1 has higher kinase activity than Clb5–Cdk1. This difference is due almost entirely to a tenfold difference in K_m values for general substrates (Table 1). K_m differences of this magnitude will have important consequences inside the cell, where these kinases are exposed to high substrate concentrations. The concentrations of Cdk1 substrates in budding yeast are difficult to estimate, because some substrates are likely to be localized to specific subcellular sites. Nevertheless, a rough estimate can be made on the basis of recent proteomic studies^{23,24}. If we combine the concentrations of the 181 CDK substrates we identified⁹, then a conservative estimate of the total concentration of CDK phosphorylation sites in the nucleus would be at least 1.0 mM. This concentration greatly exceeds the relatively low substrate K_m values (about $50 \mu\text{M}$) of Clb2–Cdk1 for general substrates. The saturation of Clb2–Cdk1 by these substrates should initially reduce the rate of phosphorylation of general substrates not involved in early S-phase events. This initial suppression of

Clb2–Cdk1 activity probably explains our observation (Fig. 3e) that phosphorylation of the general substrate Slk19 in wild-type and *clb5::CLB2* cells occurs gradually during progression through S phase, despite the fact that Slk19 is a highly efficient substrate *in vitro*. As Clb2 concentrations increase and general substrates become phosphorylated, inhibition by unphosphorylated substrates will eventually be relieved, allowing Clb2–Cdk1 to assume its role as a highly efficient kinase for the large number of mitotic substrates. In contrast, the relatively inefficient Clb5–Cdk1, because of its high K_m values for general substrates (about $500 \mu\text{M}$), might not be significantly inhibited by these proteins. This lower degree of competition from general substrates allows Clb5–Cdk1 to focus, through interactions between the hydrophobic patch and the RXL motif, on a subset of low K_m substrates whose phosphorylation is critical for S-phase initiation. The low activity of Clb5–Cdk1 towards general targets, combined with its high affinity for specific targets (see also refs 14, 25), might explain the observation that overexpressed Clb5 is unable to replace Clb2 function in blocking mitotic exit²⁶.

Our studies therefore support the existence of multiple mechanisms by which the intrinsic biochemical properties of different cyclins help to promote the correct timing of CDK substrate phosphorylation during the cell cycle. We argue that there is a fundamental advantage of this specificity over the quantitative model of cyclin function. In the quantitative model, M-phase CDK targets must have very low specificity for the kinase, because S-phase Cdk1 activity would otherwise cause their partial phosphorylation and the two phases would not be well separated. These weak substrates would then require extremely high mitotic cyclin–CDK activity to achieve significant levels of phosphorylation in M phase. However, a system employing cyclins with specific functional capacities can operate with cyclin concentrations that are lower and change less markedly, resulting in a far more efficient and robust mechanism by which the events of S and M phases are triggered in the correct order and do not overlap. □

Methods

General methods

All strains were derivatives of W303 or S288C and were grown at 30°C . Construction of epitope-tagged strains was performed as described²⁷. To construct hydrophobic patch mutations, three residues in Clb5 (Met 197, Leu 201 and Trp 204) and Clb2 (Asn 260, Leu 264 and Trp 267) were changed to alanine¹⁵. Protein extracts were prepared for immunoblotting as described⁹; tandem affinity purification (TAP) and 13Myc tags were detected by c-Myc polyclonal antibody (Santa Cruz); the three-haemagglutinin (3HA) tag was detected with the 16B12 antibody (Covance), and endogenous Clb2 was detected with polyclonal anti-Clb2 antibody (a gift from D. Kellogg). Band intensities on immunoblots were quantified by the analysis of scanned films with ImageQuant software.

Kinase purification

CDK–cyclin complexes (Cdk1–Clb2-TAP, Cdk1–Clb5-TAP, Cdk1-as1–Clb2-TAP, Cdk1-as1–Clb5-TAP, and the corresponding versions with hydrophobic patch mutations) were purified to homogeneity as described^{9,28} from yeast strains lacking *SIC1* and expressing the desired TAP-tagged cyclin under the control of the *GAL* promoter. Fin1 and Cdc6 were expressed in bacteria as His₆-tagged proteins (in pET28b) and purified by metal-affinity chromatography on a cobalt–IDA (iminodiacetic acid) column.

Kinase assays

Comparison of substrate phosphorylation with the two cyclins was performed with purified Clb–Cdk1-as1 complexes by two methods, as follows. First, glutathione S-transferase (GST)-tagged proteins²⁹ were incubated in total cell extracts with Clb–Cdk1-as1 and [γ -³²P]N⁶-(benzyl)ATP as described⁹. PhosphorImager units of substrate phosphorylation were divided by substrate amount as determined from silver-stained gels⁹. Second, for substrates that were difficult to detect in the GST-tagged library, TAP-tagged proteins²⁴ in yeast cell lysates were enriched by immobilization on magnetic IgG-coupled beads and washed; they were then subjected to kinase reactions as described above. In these cases, relative P-scores were obtained and scaled by using previously obtained P-scores for Clb2 (ref. 9). About 20 substrates – including most of the highly Clb5-specific targets – were also tested with wild-type Cdk1–Clb complexes, and the results showed that the analogue-sensitive mutation does not influence the degree of cyclin specificity.

Although our methods provide a valid approach to measuring the specificity of different cyclins for the same substrate, the significance of P-score differences for different substrates with the same cyclin remains uncertain. Substrates along the diagonal in Fig. 1b, for example, are phosphorylated at rates that vary by more than four orders of magnitude.

These differences might result from several factors, including differences in the number of phosphorylation sites on substrates, inaccuracies in our estimates of the amount of substrate in the reaction, and problems with the proteolysis and folding of GST fusion proteins. In addition, the significance of different P-scores for different proteins is difficult to assess without any knowledge of the rate at which these proteins are dephosphorylated in the cell.

Detailed kinetic analyses were performed with [γ - 32 P]ATP in a reaction mixture containing 25 mM Hepes pH 7.4, 150 mM NaCl, 10% glycerol, 1 mM EGTA, 0.1 mM ATP, 2 mM MgCl₂ and purified Clb-Cdk1 complex. Apparent K_m values were determined from initial velocities of substrate phosphorylation (up to 10% of total substrate turnover) at different substrate concentrations. Phosphorylation of protein substrates was quantified by PhosphorImager analysis of polyacrylamide gels. For peptide substrate assays, peptides were bound to phosphocellulose paper, washed with 75 mM orthophosphoric acid and quantified by counting Cerenkov radiation. For kinase assays in immunoprecipitates, cells were lysed by bead-beating in RIPA buffer, and Cdk1-Clb-3HA complexes were immunoprecipitated with 16B12 antibody and protein-G-coupled magnetic beads (DynaL Biotech). Conventional kinase assays with [γ - 32 P]ATP and histone H1 were then performed. Peptide substrates were obtained from Promega and Sigma.

Received 25 October; accepted 20 December 2004; doi:10.1038/nature03329.

1. Miller, M. E. & Cross, F. R. Cyclin specificity: how many wheels do you need on a unicycle? *J. Cell Sci.* **114**, 1811–1820 (2001).
2. Roberts, J. M. Evolving ideas about cyclins. *Cell* **98**, 129–132 (1999).
3. Stern, B. & Nurse, P. A quantitative model for the cdc2 control of S phase and mitosis in fission yeast. *Trends Genet.* **12**, 345–350 (1996).
4. Fisher, D. L. & Nurse, P. A single fission yeast mitotic cyclin B p34^{cdc2} kinase promotes both S-phase and mitosis in the absence of G1 cyclins. *EMBO J.* **15**, 850–860 (1996).
5. Moore, J. D., Kirk, J. A. & Hunt, T. Unmasking the S-phase-promoting potential of cyclin B1. *Science* **300**, 987–990 (2003).
6. Cross, F. R., Yuste-Rojas, M., Gray, S. & Jacobson, M. D. Specialization and targeting of B-type cyclins. *Mol. Cell* **4**, 11–19 (1999).
7. Donaldson, A. D. The yeast mitotic cyclin Clb2 cannot substitute for S phase cyclins in replication origin firing. *EMBO Rep.* **1**, 507–512 (2000).
8. Peeper, D. S. et al. A- and B-type cyclins differentially modulate substrate specificity of cyclin-cdk complexes. *EMBO J.* **12**, 1947–1954 (1993).
9. Ubersax, J. A. et al. Targets of the cyclin-dependent kinase Cdk1. *Nature* **425**, 859–864 (2003).
10. Brown, N. R., Noble, M. E., Endicott, J. A. & Johnson, L. N. The structural basis for specificity of substrate and recruitment peptides for cyclin-dependent kinases. *Nature Cell Biol.* **1**, 438–443 (1999).
11. Adams, P. D. et al. Identification of a cyclin-cdk2 recognition motif present in substrates and p21-like cyclin-dependent kinase inhibitors. *Mol. Cell Biol.* **16**, 6623–6633 (1996).
12. Schulman, B. A., Lindstrom, D. L. & Harlow, E. Substrate recruitment to cyclin-dependent kinase 2 by a multipurpose docking site on cyclin A. *Proc. Natl Acad. Sci. USA* **96**, 10453–10458 (1998).
13. Takeda, D. Y., Wohlschlegel, J. A. & Dutta, A. A bipartite substrate recognition motif for cyclin-dependent kinases. *J. Biol. Chem.* **276**, 1993–1997 (2001).
14. Wilmes, G. M. et al. Interaction of the S-phase cyclin Clb5 with an 'RXL' docking sequence in the initiator protein Orc6 provides an origin-localized replication control switch. *Genes Dev.* **18**, 981–991 (2004).
15. Cross, F. R. & Jacobson, M. D. Conservation and function of a potential substrate-binding domain in the yeast Clb5 B-type cyclin. *Mol. Cell Biol.* **20**, 4782–4790 (2000).
16. van Hemert, M. J. et al. The *Saccharomyces cerevisiae* Fin1 protein forms cell cycle-specific filaments between spindle pole bodies. *Proc. Natl Acad. Sci. USA* **99**, 5390–5393 (2002).
17. Goldberger, A. & Koshland, D. E. Jr An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl Acad. Sci. USA* **78**, 6840–6844 (1981).
18. Masumoto, H., Muramatsu, S., Kamimura, Y. & Araki, H. S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature* **415**, 651–655 (2002).
19. Peters, J. M. The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol. Cell* **9**, 931–943 (2002).
20. Diffley, J. F. Regulation of early events in chromosome replication. *Curr. Biol.* **14**, R778–R786 (2004).
21. Yeong, F. M., Lim, H. H., Padmashree, C. G. & Surana, U. Exit from mitosis in budding yeast: biphasic inactivation of the Cdc28-Clb2 mitotic kinase and the role of Cdc20. *Mol. Cell* **5**, 501–511 (2000).
22. Wäsch, R. & Cross, F. APC-dependent proteolysis of the mitotic cyclin Clb2 is essential for mitotic exit. *Nature* **418**, 556–562 (2002).
23. Cross, F. R., Archambault, V., Miller, M. & Klovstad, M. Testing a mathematical model of the yeast cell cycle. *Mol. Biol. Cell* **13**, 52–70 (2002).
24. Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
25. Archambault, V. et al. Targeted proteomic study of the cyclin-cdk module. *Mol. Cell* **14**, 699–711 (2004).
26. Jacobson, M. D., Gray, S., Yuste-Rojas, M. & Cross, F. R. Testing cyclin specificity in the exit from mitosis. *Mol. Cell Biol.* **20**, 4483–4493 (2000).
27. Longtine, M. S. et al. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
28. Puig, O. et al. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
29. Martzen, M. R. et al. A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank E. Woodbury for Fin1 reagents and assistance; J. Ubersax for numerous reagents and discussions; K. Shokat for N⁶-(benzyl)ADP; E. Phizicky for the GSTopen reading frame library; E. K. O'Shea and J. S. Weissman for TAP-tagged strains; and C. Carroll, L. Holt, M. Matyskiela, D. Randle and other members of the Morgan laboratory for discussions and comments on the manuscript. This work was supported by funding from the National Institute of General Medical Sciences.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.O.M. (dmorgan@cgl.ucsf.edu).

Defective DNA single-strand break repair in spinocerebellar ataxia with axonal neuropathy-1

Sherif F. El-Khamisy¹, Gulam M. Saifi², Michael Weinfeld³, Fredrik Johansson⁴, Thomas Helleday^{4,5}, James R. Lupski² & Keith W. Caldecott¹

¹Genome Damage and Stability Centre, University of Sussex, Science Park Road, Falmer, Brighton BN1 9RQ, UK

²Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Room 604B, Houston, Texas 77030, USA

³Cross Cancer Institute, Edmonton, Alberta T6G1Z2, Canada

⁴Department of Genetics Microbiology and Toxicology, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

⁵The Institute for Cancer Studies, University of Sheffield, Medical School, Beech Hill Road, Sheffield S10 2RX, UK

Spinocerebellar ataxia with axonal neuropathy-1 (SCAN1) is a neurodegenerative disease that results from mutation of tyrosyl phosphodiesterase 1 (TDP1)¹. In lower eukaryotes, Tdp1 removes topoisomerase 1 (top1) peptide from DNA termini during the repair of double-strand breaks created by collision of replication forks with top1 cleavage complexes in proliferating cells^{2–4}. Although TDP1 most probably fulfils a similar function in human cells, this role is unlikely to account for the clinical phenotype of SCAN1, which is associated with progressive degeneration of post-mitotic neurons. In addition, this role is redundant in lower eukaryotes, and Tdp1 mutations alone confer little phenotype^{4–7}. Moreover, defects in processing or preventing double-strand breaks during DNA replication are most probably associated with increased genetic instability and cancer, phenotypes not observed in SCAN1 (ref. 8). Here we show that in human cells TDP1 is required for repair of chromosomal single-strand breaks arising independently of DNA replication from abortive top1 activity or oxidative stress. We report that TDP1 is sequestered into multi-protein single-strand break repair (SSBR) complexes by direct interaction with DNA ligase III α and that these complexes are catalytically inactive in SCAN1 cells. These data identify a defect in SSBR in a neurodegenerative disease, and implicate this process in the maintenance of genetic integrity in post-mitotic neurons.

To investigate the molecular basis of SCAN1, normal and SCAN1 lymphoblastoid cells were compared for levels of DNA breakage during a 1-h treatment with camptothecin (CPT). CPT increases the half-life of cleavage complex intermediates of top1 activity and so increases the likelihood of their conversion into strand breaks⁹, which can be measured with the alkaline comet assay¹⁰. Normal cells accumulated low levels of breakage during the first 20 min of treatment with CPT, which then typically fell to near-background levels by the end of treatment (Fig. 1a). In contrast, SCAN1 cells

continued to accumulate breaks, resulting in about eightfold more breaks than normal cells by the end of drug treatment (about 10 Gy equivalent or 10,000 total breaks per cell). Because these data did not reflect increased apoptosis (data not shown) or altered formation or reversal of top1 cleavage complexes (Fig. 1b), we concluded that SCAN1 cells are unable to repair CPT-induced strand breaks rapidly. Consistent with this was our observation that CPT-induced breaks in SCAN1 cells failed to decline even if these cells were subsequently incubated in drug-free medium for 1 h (Fig. 1c). This was in contrast to normal cells, in which strand breaks fell by 70–80% despite having been treated with a tenfold higher

dose of CPT to induce similar levels of strand breakage. Similar results were observed if alkaline unwinding was employed to measure strand breakage (Supplementary Fig. 1).

To examine whether strand breaks accumulate in a replication-independent manner in SCAN1 cells, and might therefore arise in post-mitotic neurons, we inhibited DNA replication to about 6% of control levels by preincubation with mimosine. Mimosine ablated the accumulation of strand breaks in normal cells, confirming that these were entirely dependent on replication (Fig. 1d, e). In striking contrast, preincubation with mimosine reduced the accumulation of breaks in SCAN1 cells by only half. The remaining strand breaks

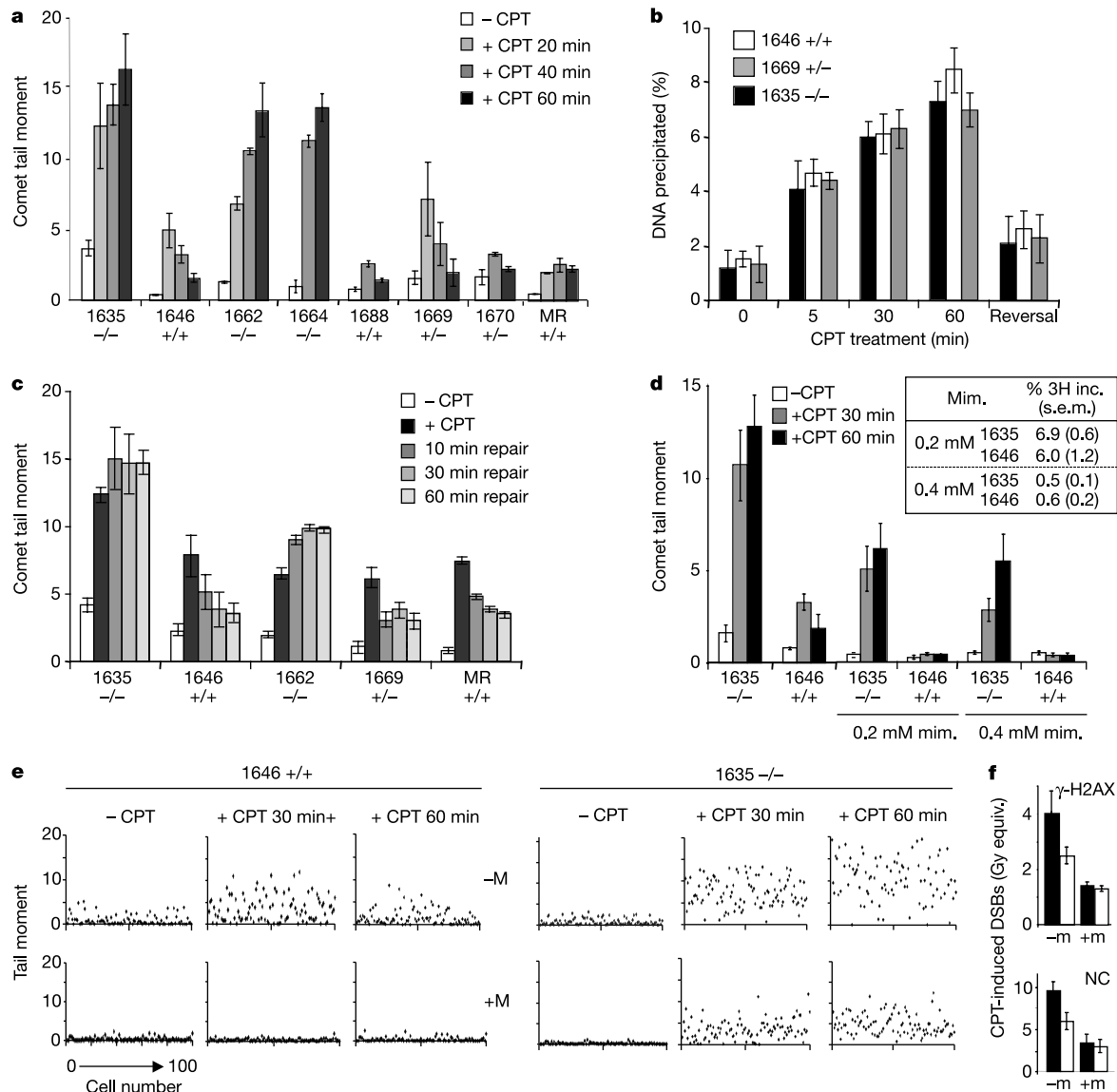


Figure 1 Defective repair of replication-independent top1-SSBs in SCAN1. **a**, Strand breaks were quantified as comet tail moments in mock-treated or CPT-treated (14 μ M, 0–60 min) cells by alkaline comet assays. +/+, +/- and -/- denote cells wild-type, heterozygous and homozygous for TDP1 mutation, respectively¹. Numbers identify individual cell lines. **b**, Top1 cleavage complexes quantified by protein–DNA precipitation assays after treatment with CPT (14 μ M) for 0–60 min, or after 60 min followed by a subsequent 30 min in drug-free medium (reversal). **c**, Strand breaks quantified in untreated cells (–CPT), cells treated with 14 μ M (SCAN1) or 140 μ M (normal) CPT for 30 min (+CPT), or cells treated for 30 min and then incubated for 10–60 min in drug-free medium. **d**, Strand breaks quantified in mock-treated or CPT-treated (14 μ M) cells with or

without preincubation with 0.2 or 0.4 mM mimosine (mim.). Residual DNA replication levels (percentage of control) are shown in the inset. **e**, Representative scatter plots of 100 individual comet tail moments per sample (cells numbered from 0 to 100) for mock-treated or CPT-treated cells with or without preincubation with 0.2 mM mimosine (+/–m). **f**, CPT-induced DSBs (14 μ M, 1 h) measured by quantitative γ -H2AX immunofluorescence (top) and neutral comet assays (NC, bottom) in normal (BAB1646, open bars) or SCAN1 (BAB1635, filled bars) cells with or without preincubation with 0.2 mM mimosine (+/–m). All histograms and replication data are means $\pm$ s.e.m. of at least three independent experiments.

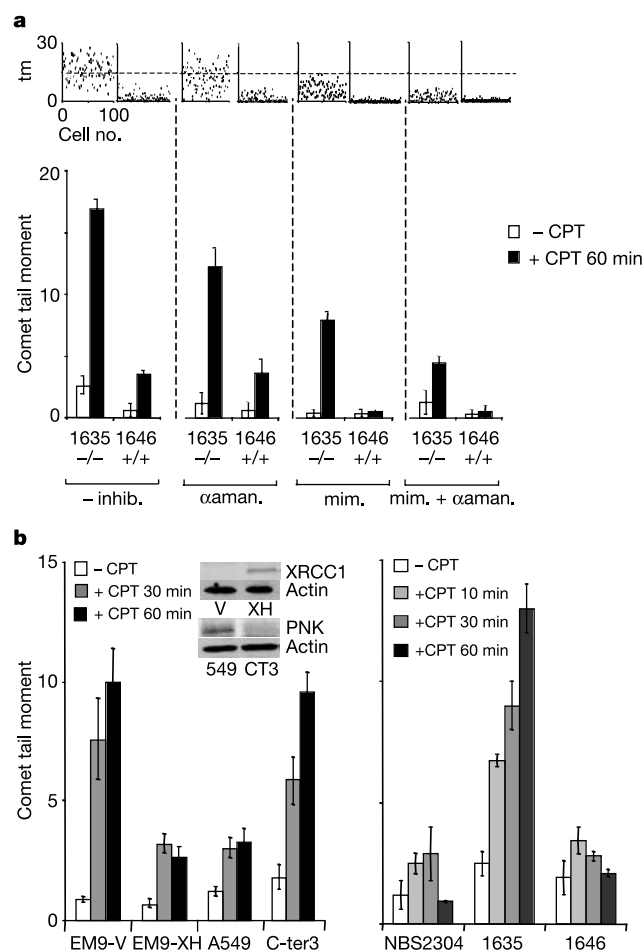


Figure 2 Defective repair of top1-SSBs in other SSBR mutant cells. **a**, Strand breaks were quantified in mock-treated or CPT-treated (14 μ M, 1 h) normal or SCAN1 cells preincubated with 5 μ g ml⁻¹ α -amanitin (+ α aman.), 0.2 mM mimosine (mim.), both (mim. + α aman.), or none (-inhib.) for 16 h. Residual transcription after α -amanitin was $7.1 \pm 1.1\%$ (mean $\pm$ s.e.m.). Representative scatter plots of 100 individual comet tail moments (tm) per sample (cells numbered from 0 to 100) are shown for each CPT-treated sample above the relevant filled histogram. Horizontal dotted line denotes middle of tail moment scale. **b**, Strand breaks were quantified in mock-treated or CPT-treated (14 μ M, 10–60 min) *XRCC1*-mutant EM9 Chinese hamster ovary cells harbouring empty expression vector (EM9-V) or human *XRCC1* cDNA (EM9-XH), or in human A549 lung carcinoma cells or a derivative of A549 cells with stable PNK knockdown (C-ter3), or in normal (BAB 1646), SCAN1 (BAB 1635) or Lig4 syndrome (NBS2304) LCLs. The inset shows *XRCC1*, *PNK* and actin protein levels in relevant cell lines. All histograms are means $\pm$ s.e.m. of at least three independent experiments.

did not result from residual replication, because increasing the concentration of mimosine reduced replication a further tenfold, to 0.5% of control levels, but did not further reduce strand breakage. Thus, about half of the breaks that accumulate in CPT-treated SCAN1 cells arise independently of replication, and so are a previously undescribed type of break not detected in normal cells.

To examine whether the breaks that accumulate in SCAN1 cells were single-strand breaks (SSBs) or double-strand breaks (DSBs), we conducted quantitative γ -H2AX immunostaining and neutral comet assays, two techniques for measuring DSBs. If the CPT-induced breaks in SCAN1 cells were composed entirely of DSBs, this would be equivalent to a γ -ray dose of 60–120 Gy (2,500–5,000 DSBs). However, neither assay revealed a measurable difference between the level of DSBs in non-replicating normal and SCAN1

cells (Fig. 1f). Consequently, because both assays can detect as little as 1–2 Gy equivalents of DSBs, more than 98% of the breaks that accumulate in non-replicating SCAN1 cells are SSBs. In cycling SCAN1 cells an increase in DSBs of 2–4 Gy equivalents per cell was observed, compared with normal cells. This level is equivalent to less than 4% of the total breaks detected in proliferating SCAN1 cells and is consistent with a number of unrepaired SSBs encountering DNA replication during the 1-h treatment with CPT.

With regard to the source of SSBs in SCAN1 cells, replication-independent top1-associated SSBs (top1-SSBs) can arise by collision of cleavage complexes with RNA polymerases or by close proximity to other DNA adducts^{11–13}. Consistent with this was our observation that the accumulation of replication-independent SSBs in CPT-treated SCAN1 cells was reduced by almost half by pre-inhibition with the RNA polymerase inhibitor α -amanitin (Fig. 2a). In lower eukaryotes, top1-SSBs are presumably repaired once they have been converted into DSBs during DNA replication. However, to explain our data, we proposed that human cells possess a novel TDP1-dependent SSBR process that can rapidly repair most top1-SSBs before they are converted into DSBs, and that SCAN1 cells lack this process. If true, CPT-induced DNA strand breaks should also accumulate in mutant cell lines with established defects in SSBR. Indeed, both EM9 Chinese hamster ovary cells lacking *XRCC1*–DNA ligase III α heterodimer¹⁴ and human A549 cells (C-ter3)¹⁵ harbouring depleted levels of polynucleotide kinase (PNK) accumulated CPT-induced strand breaks to a level comparable with those in SCAN1 cells (Fig. 2b). *XRCC1* is a scaffold protein that assembles SSBR protein complexes¹⁶ and PNK is a DNA kinase/phosphatase that processes damaged DNA ends^{15,17,18}. In contrast, CPT-induced DNA strand breaks did not accumulate above normal levels in lymphoblastoid cells from an individual with Lig4 syndrome, in which the non-homologous end-joining pathway for DSB repair is defective¹⁹ (Fig. 2b, right panel). We therefore conclude that normal human cells possess a novel TDP1-dependent SSBR process that rapidly repairs top1-SSBs, and that this process is absent in SCAN1.

Yeast two-hybrid experiments revealed that TDP1 associates with the SSBR machinery by direct interaction with DNA ligase III α (Lig3 α), a partner of the critical SSBR protein *XRCC1* (Fig. 3a)^{14,20–22}. This might account for the observation that TDP1 co-immunoprecipitates with *XRCC1* (ref. 23). Strikingly, TDP1 immunoprecipitate recovered from normal cell extract, but notably not SCAN1 cell extract, was able to repair tyrosyl-associated oligonucleotide substrates that mimic top1-SSBs *in vitro*, indicating that TDP1 might be a component of a complex containing all enzymes necessary for TDP1-dependent SSBR (Fig. 3b). The predicted steps of this process involve the removal of tyrosine (Y) from the 3' terminus of the ³²P-labelled 18-mer by TDP1, removal of the resulting 3'-phosphate (P) by PNK, and finally ligation of the nick by Lig3 α (Fig. 3b, right panel). This model agrees with the inability of mutant cells lacking these activities to rapidly repair SSBs induced by CPT (Fig. 2b), with the observed presence of TDP1, *XRCC1*, PNK and Lig3 α in TDP1 immunoprecipitates from normal cell extract (Fig. 3c, lane 4), and with the reconstituted repair of tyrosyl-SSBs by recombinant human TDP1, PNK and Lig3 α (Fig. 3d). Although *XRCC1* was not required for TDP1-dependent SSBR reactions *in vitro*, it is most probably required for assembly of the TDP1 SSBR complexes in cells¹⁶. *XRCC1*, PNK and Lig3 α were also present in TDP1 immunoprecipitates from SCAN1 extract (Fig. 3c, lane 3), indicating that the SSBR defect in SCAN1 might reflect TDP1 inactivity rather than an absence of the SSBR complex, a notion confirmed by the restoration of SSBR capacity to SCAN1 TDP1 immunoprecipitates by recombinant human TDP1 (Fig. 3e, lane 5).

Next, we examined whether TDP1-dependent SSBR is also required for the repair of SSBs induced by oxidative damage, because neurons exhibit high levels of oxidative stress and oxidative

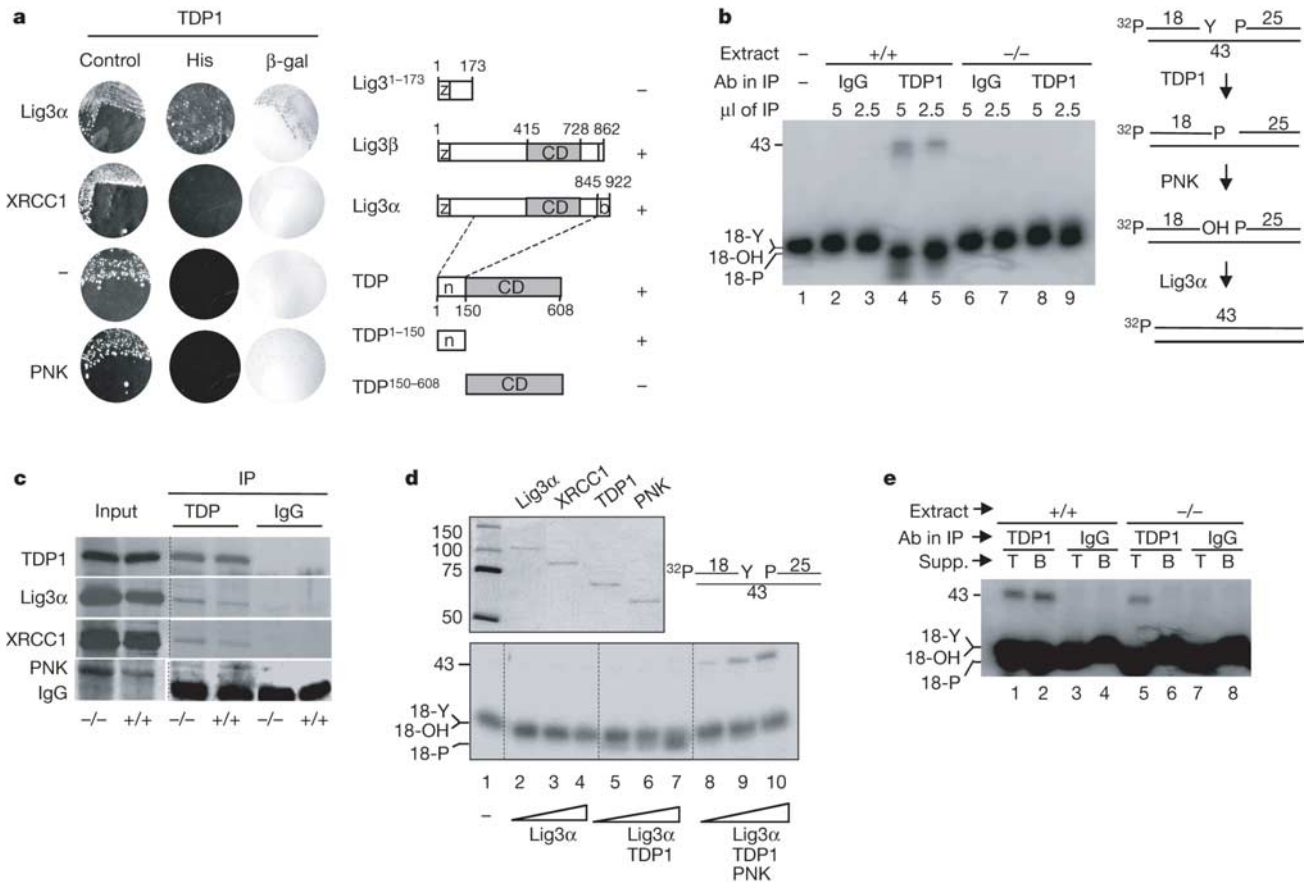


Figure 3 TDP1 interacts with DNA ligase III α and is a component of SSBR multi-protein complexes. **a**, Left, yeast Y190 cells harbouring pGBKT7-TDP1 and pACT plasmid encoding either Lig3 α , XRCC1, PNK or empty PACT were plated on control plates and plates appropriate for detecting the activation of *His3* (His) and β -galactosidase (β -gal) reporter genes. Right, summary of two-hybrid experiments. The presence or absence (+/-) of interaction between TDP1 and Lig3 α , Lig3 β or Lig3 α (top three bars), and the presence or absence of interaction between Lig3 α and TDP1, TDP1 $^{1-150}$, or TDP1 $^{150-608}$ (bottom three bars) are shown. Dotted lines denote interacting region. Z, CD, n and b denote zinc finger, catalytic, amino-terminal and BRCT (BRCA1 C-terminal) domains. **b**, 43-mer substrate containing a nick with 3'-tyrosyl terminus (Y) was incubated for 1 h at 37 °C with 2.5–5.0 μ l of TDP1 immunoprecipitate (IP) or IgG immunoprecipitate from normal (BAB1646; +/+) or SCAN1 (BAB1635; -/-) cell extract (2.25 $\times$ more SCAN1 extract protein was used in IP to compensate for lower TDP1

levels). Reaction products were separated by denaturing PAGE. Positions of 43-mer product (43), 32 P-labelled 18-mer substrate (18-Y), and repair intermediates (18-P and 18-OH) are indicated on the left of the gel and their predicted origin on the right. **c**, Input extracts and TDP1 and IgG immunoprecipitates employed above were immunoblotted for the indicated proteins. **d**, Top, 0.1 μ g of the indicated purified recombinant human polypeptides was fractionated by SDS-PAGE and stained with Coomassie blue. Bottom, the 43-mer duplex substrate (inset) was incubated in the absence (-) or presence of 1 nM (lanes 2, 5 and 8), 2 nM (lanes 3, 6 and 9) or 4 nM (lanes 4, 7 and 10) of each of the indicated purified recombinant human proteins for 1 h at 37 °C. **e**, 43-mer duplex substrate was incubated with TDP1 or IgG immunoprecipitates (IP) as described in **b** in the additional presence (supp.) of 7 nM recombinant human TDP1 (T) or BSA (B).

DNA damage can 'trap' cleavage complexes²⁴. In addition, oxidative stress can induce SSBs directly, some of which might be substrates for TDP1 (refs 5, 25). Whereas there was no significant difference between SCAN1 and normal cells in the level of strand breaks induced by H₂O₂, significantly greater levels persisted in SCAN1 cells after either 30 min or 1 h repair periods in H₂O₂-free medium (Fig. 4a). This difference was apparent even 10 min after the removal of H₂O₂, which is consistent with the rapidity of SSBR processes in normal cells (data not shown). Because more than 99% of strand breaks initially induced by H₂O₂ are SSBs²⁶, these data demonstrate that oxidative SSBs accumulate in SCAN1 cells. Next we considered how the defect in SSBR in SCAN1 cells might affect cell function. Because SSBs can block transcription^{27,28}, we examined whether transcriptional inhibition was greater and/or more prolonged in SCAN1 cells. Indeed, although levels of transcription were reduced to similar levels in normal and SCAN1 cells during a 1-h incubation

with CPT (Fig. 4b), only in normal cells did transcription recover during a subsequent 19-h incubation in drug-free medium (Fig. 4c). In SCAN1 cells, transcription failed to recover during this period at any concentration of CPT employed.

Why might the pathology of SCAN1 be restricted to neurons? One possibility is that the elevated oxidative stress in post-mitotic neurons creates a particularly high level of SSBs. In addition, the limited regenerative capacity of the nervous system might render this tissue particularly susceptible to cell loss, whereas cell loss from tissues with greater regenerative capacity might be better tolerated. Another possibility might be that unrepaired SSBs can be processed by homologous recombination (HR), in proliferating cells, even in the absence of TDP1, once they have been converted into DSBs during DNA replication, as in budding yeast. Consistent with this idea was our observation that HR-defective *XRCC3*^{-/-} chicken DT40 cells accumulated high levels of replication-dependent strand

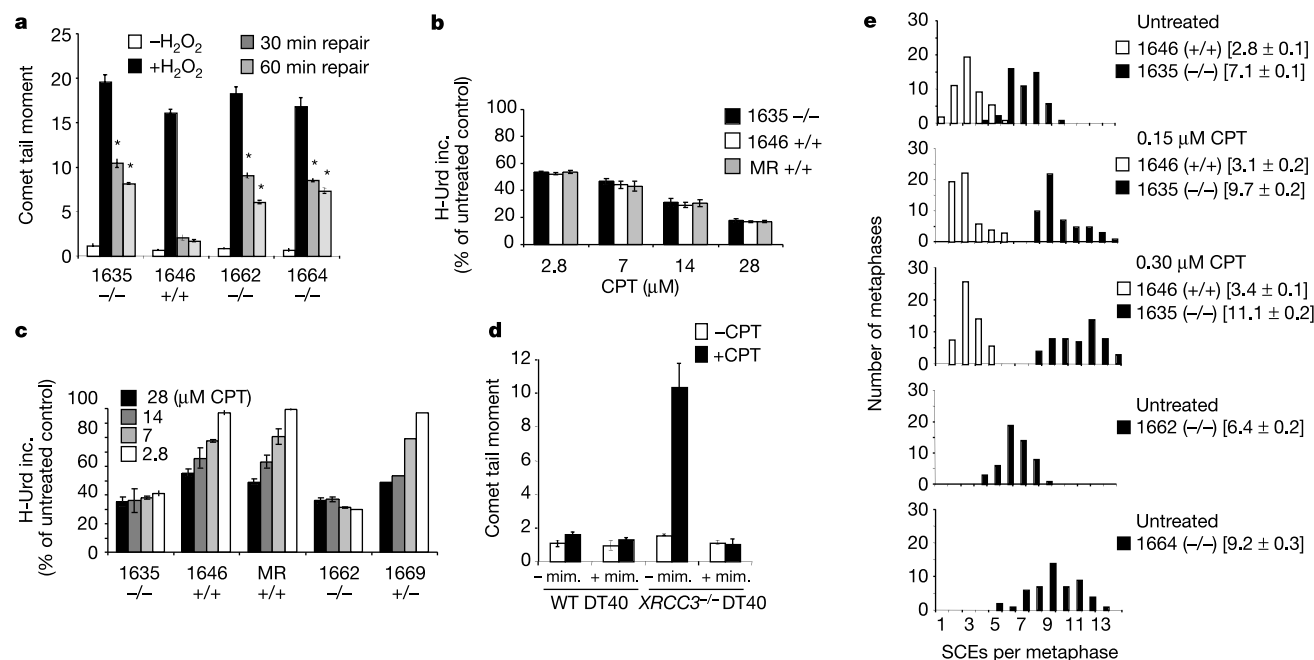


Figure 4 Repair of oxidative SSBs, CPT-induced transcriptional inhibition, and sister chromatid exchanges in SCAN1 cells. **a**, Strand breaks were quantified in the indicated cells after mock treatment or treatment with H_2O_2 (150 μM , 10 min on ice), and after repair incubations for 30–60 min in drug-free medium. Asterisks denote statistically significant ($P < 0.01$; t -test) differences between SCAN1 and normal H_2O_2 histograms at the indicated time points. **b**, Levels of transcription (relative to untreated control) in the indicated cells during a 1-h 3H -uridine pulse label in the presence of 2.8–28 μM CPT. **c**, Levels of transcription (relative to untreated control) during a 1-h 3H -uridine pulse label

after a 19-h recovery period in drug-free medium in cells first treated as in **b**. **d**, Strand breaks were quantified in mock-treated (–CPT) or CPT-treated (+CPT; 14 μM for 1 h) wild-type or $XRCC3^{-/-}$ DT40 cells preincubated or not with 0.2 mM mimosine (mim.), by alkaline comet assays. **e**, SCE frequencies in the indicated untreated or CPT-treated (150–300 nM for 1 h) normal (open bars) or SCAN1 (filled bars) cells. SCEs per cell (means $\pm$ s.e.m.) are indicated in square brackets. All other histograms in this figure are means $\pm$ s.e.m. of at least three independent experiments.

breaks during treatment with CPT, confirming that HR repairs replication-associated breaks in higher eukaryotes (Fig. 4d). Moreover, HR is indeed elevated in SCAN1 cells, as indicated by an increased frequency of ‘spontaneous’ sister chromatid exchanges (SCEs) and by an additional increase in these events in SCAN1 cells after exposure to concentrations of CPT too low to trigger SCEs in normal cells (Fig. 4e). HR can therefore fulfil a back-up role in the absence of TDP-dependent SSBR, allowing proliferating SCAN1 cells to tolerate low levels of unrepaired SSBs.

In conclusion, we have identified a novel TDP1-dependent SSBR process that is required in the repair of chromosomal SSBs created by abortive top1 activity and oxidative stress, and we demonstrate a defect in this process in a neurodegenerative disease. □

Methods

Lymphoblastoid cell lines

Normal (BAB1646, BAB1688, BAB1669, BAB1670) and SCAN1 (BAB1635, BAB1662, BAB1664) lymphoblastoid cell lines (LCLs) were provided by G.M.S and J.R.L. MR is an unrelated normal LCL provided by M. Taylor.

Yeast two-hybrid assay

Y190 cells harbouring pGBKT7-TDP1 and either pACT-Lig3 α , pACT-XRCC1, pACT-PNK or empty pACT were plated on control minimal medium lacking Leu and Trp to select for plasmids, and medium additionally lacking His and containing 25 mM 3-aminotriazole to select for the *His3* reporter activation. Filter-lifts from control plates were employed for measuring β -galactosidase reporter activation.

Comet assays

Alkaline comet assays measure SSBs and DSBs by their ability to increase DNA migration out of the nucleus during electrophoresis¹⁰. Neutral assays measure DSBs. Strand breakage is quantified as the tail moment, an arbitrary unit reflecting the product of the amount of DNA that leaves the nucleus and the distance migrated. Treated cells were mixed with low-gelling-temperature agarose and layered on agarose-coated slides on ice. For alkaline assays, lysis was in 2.5 M NaCl, 10 mM Tris-HCl pH 7.0, 0.1 M EDTA, 1% Triton X-100,

1% dimethyl sulphoxide (DMSO; pH 10) for 1 h at 4 °C. Electrophoresis was in 50 mM NaOH, 1 mM EDTA, 1% DMSO at 25 V (0.6 V cm^{-1}) for 25 min. For neutral assays, lysis was in 2.5 M NaCl, 10 mM Tris-HCl pH 7.0, 0.1 M EDTA, 0.5% Triton X-100, 3% DMSO, 1% *N*-lauroylsarcosine (pH 9.5) for 90 min at 4 °C, and electrophoresis was in 0.3 M sodium acetate, 100 mM Tris-HCl, 1% DMSO (pH 8.3) for 1 h. Slides were neutralized in 0.4 M Tris-HCl pH 7.0 and DNA was stained with Sybr Green I. Mean tail moments were quantified for 100 cells per sample in each experiment, with Comet Assay III software (Perceptive Instruments). Histograms are the average mean tail moment per sample of at least three experiments (reported as means $\pm$ s.e.m.).

Replication and transcription

^{14}C -thymidine-labelled cells treated as indicated were pulse-labelled with 20 $\mu Ci ml^{-1}$ 3H -thymidine for 1 h, or with 10 $\mu Ci ml^{-1}$ 3H -uridine for 4 h, to measure DNA or RNA synthesis, respectively. Trichloroacetic acid-precipitable counts were measured by liquid scintillation and 3H counts were normalized by using ^{14}C counts. To measure the impact of CPT on transcription, CPT was included during the 1 h of 3H -uridine pulse labelling. To measure transcription recovery, CPT-treated cells were subsequently incubated in drug-free medium for 19 h and pulse-labelled with 3H -uridine for 1 h.

Cleavage complexes

To measure cleavage complexes, 2.5×10^5 3H -thymidine-labelled cells were treated as indicated and then lysed in 1.25% SDS, 5 mM EDTA, 0.4 mg ml^{-1} salmon sperm DNA, and DNA was sheared by passage (20 times) through a 0.4-mm diameter needle. KCl was added to 65 mM and protein–DNA complexes were precipitated on ice for 10 min, pelleted in a Microfuge for 10 min at 4 °C, and washed by resuspension in 10 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM EDTA, 0.1 mg ml^{-1} salmon sperm DNA at 60 °C for 10 min. Precipitation and washing were repeated twice and the fraction of DNA present in precipitates, relative to total DNA present before precipitation, was quantified by liquid scintillation.

Recombinant proteins, immunoprecipitations and repair assays

Histidine-tagged recombinant human TDP1, XRCC1, PNK and Lig3 α were expressed in BL21(DE3) cells and purified by metal-chelate chromatography. Lig3 α and XRCC1 were further purified by gel filtration, and TDP1 was purified by ion-exchange chromatography. For immunoprecipitation, 6×10^6 cells were lysed in 20 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 mM EGTA, 100 mM NaCl, 1% Triton X-100, containing protease inhibitors (Roche) and incubated with anti-TDP1 polyclonal antibodies (Abcam) or rabbit IgG (Dako) for 3 h on ice, followed by a further 90 min in the additional presence of

Protein G-Sepharose beads (Amersham). Washed immunocomplexes were fractionated by SDS-PAGE for immunoblotting with anti-TDP1, anti-Lig3 α (TL25), anti-XRCC1 (SeroTec AHP428) or anti-PNK (CytoStore) polyclonal antibodies or were resuspended in 20 mM HEPES, 0.1 M NaCl, 1 mM dithiothreitol, 10% glycerol for repair assays. Sequences of the 18-mer, 25-mer and 43-mer oligonucleotides for repair assays were 5'-TCCGTTGAAGCCTGCTTT-3', 5'-GACATACTAACTTGAGCGAAACGGT-3' and 5'-TAGGCAACTTCGGACGAACTGTATGATTGAACCTCGCTTTGGCC-3', respectively. An 18-mer containing a tyrosine 3' terminus (5'-TCCGTTGAAGCCTGCTTT-Tyr-3') was provided by H. Nash and has been described previously⁴. 5'-phosphorylated oligonucleotides were annealed in appropriate combinations and incubated at 25 nM in 10 μ l total volume with recombinant proteins or with anti-TDP1 or IgG immunoprecipitate recovered from 0.4 mg (normal cells) or 0.9 mg (SCAN1 cells) of cell extract in 25 mM HEPES pH 8, 130 mM KCl, 1 mM dithiothreitol, 10 mM MgCl₂, 1 mM ATP. Precipitate from about twice as much SCAN1 cell extract was employed to compensate for the lower concentrations of TDP1 in this cell line. Reactions were terminated with loading buffer and fractionated by denaturing PAGE.

γ -H2AX and SCEs

Fixed and permeabilized cells were stained with anti- γ -H2AX monoclonal antibody (clone JBW301; Upstate) followed by fluorescein isothiocyanate-conjugated anti-mouse IgG (Dako). Nuclei counterstained with 4,6-diamidino-2-phenylindole were analysed with a Zeiss Axioplan-2 fluorescence microscope and the intensity of green fluorescence was quantified with SimplePci imaging software. SCEs were measured with the approach of Perry and Wolff²⁹, and performed essentially as described³⁰.

Received 2 September; accepted 22 December 2004; doi:10.1038/nature03314.

1. Takashima, H. *et al.* Mutation of TDP1, encoding a topoisomerase I-dependent DNA damage repair enzyme, in spinocerebellar ataxia with axonal neuropathy. *Nature Genet.* **32**, 267–272 (2002).
2. Pouliot, J. J., Yao, K. C., Robertson, C. A. & Nash, H. A. Yeast gene for a Tyr-DNA phosphodiesterase that repairs topoisomerase I complexes. *Science* **286**, 552–555 (1999).
3. Li, T. K. & Liu, L. F. Tumor cell death induced by topoisomerase-targeting drugs. *Annu. Rev. Pharmacol. Toxicol.* **41**, 53–77 (2001).
4. Pouliot, J. J., Robertson, C. A. & Nash, H. A. Pathways for repair of topoisomerase I covalent complexes in *Saccharomyces cerevisiae*. *Genes Cells* **6**, 677–687 (2001).
5. Liu, C., Pouliot, J. J. & Nash, H. A. Repair of topoisomerase I covalent complexes in the absence of the tyrosyl-DNA phosphodiesterase Tdp1. *Proc. Natl Acad. Sci. USA* **99**, 14970–14975 (2002).
6. Vance, J. R. & Wilson, T. E. Yeast Tdp1 and Rad1-Rad10 function as redundant pathways for repairing Top1 replicative damage. *Proc. Natl Acad. Sci. USA* **99**, 13669–13674 (2002).
7. Liu, C., Pouliot, J. J. & Nash, H. A. The role of TDP1 from budding yeast in the repair of DNA damage. *DNA Repair (Amst.)* **3**, 593–601 (2004).
8. Shivji, M. K. & Venkitaraman, A. R. DNA recombination, chromosomal stability and carcinogenesis: insights into the role of BRCA2. *DNA Repair (Amst.)* **3**, 835–843 (2004).
9. Hsiang, Y. H., Hertzberg, R., Hecht, S. & Liu, L. F. Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I. *J. Biol. Chem.* **260**, 14873–14878 (1985).
10. Fairbairn, D. W., Olive, P. L. & O'Neill, K. L. The comet assay: a comprehensive review. *Mutat. Res.* **339**, 37–59 (1995).
11. Bendixen, C., Thomsen, B., Alsner, J. & Westergaard, O. Camptothecin-stabilized topoisomerase I-DNA adducts cause premature termination of transcription. *Biochemistry* **29**, 5613–5619 (1990).
12. Pommier, Y. *et al.* Repair of and checkpoint response to topoisomerase I-mediated DNA damage. *Mutat. Res.* **532**, 173–203 (2003).
13. Wu, J. & Liu, L. F. Processing of topoisomerase I cleavable complexes into DNA damage by transcription. *Nucleic Acids Res.* **25**, 4181–4186 (1997).
14. Caldecott, K. W., Tucker, J. D., Stanker, L. H. & Thompson, L. H. Characterization of the XRCC1-DNA ligase III complex *in vitro* and its absence from mutant hamster cells. *Nucleic Acids Res.* **23**, 4836–4843 (1995).
15. Rasouli-Nia, A., Karimi-Busheri, F. & Weinfeld, M. Stable down-regulation of human polynucleotide kinase enhances spontaneous mutation frequency and sensitizes cells to genotoxic agents. *Proc. Natl Acad. Sci. USA* **101**, 6905–6910 (2004).
16. Caldecott, K. W. XRCC1 and DNA strand break repair. *DNA Repair (Amst.)* **2**, 955–969 (2003).
17. Loizou, J. I. *et al.* The protein kinase CK2 facilitates repair of chromosomal DNA single-strand breaks. *Cell* **117**, 17–28 (2004).
18. Whitehouse, C. J. *et al.* XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* **104**, 1–11 (2001).
19. O'Driscoll, M. *et al.* DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol. Cell* **8**, 1175–1185 (2001).
20. Caldecott, K. W., Mckeown, C. K., Tucker, J. D., Ljungquist, S. & Thompson, L. H. An interaction between the mammalian DNA repair protein XRCC1 and DNA ligase III. *Mol. Cell. Biol.* **14**, 68–76 (1994).
21. Moore, D. J., Taylor, R. M., Clements, P. & Caldecott, K. W. Mutation of a BRCT domain selectively disrupts DNA single-strand break repair in noncycling Chinese hamster ovary cells. *Proc. Natl Acad. Sci. USA* **97**, 13649–13654 (2000).
22. Taylor, R. M., Moore, D. J., Whitehouse, J., Johnson, P. & Caldecott, K. W. A cell cycle-specific requirement for the XRCC1 BRCT II domain during mammalian DNA strand break repair. *Mol. Cell. Biol.* **20**, 735–740 (2000).
23. Plo, I. *et al.* Association of XRCC1 and tyrosyl DNA phosphodiesterase (Tdp1) for the repair of topoisomerase I-mediated DNA lesions. *DNA Repair (Amst.)* **2**, 1087–1100 (2003).
24. Pourquier, P. *et al.* Induction of reversible complexes between eukaryotic DNA topoisomerase I and DNA-containing oxidative base damages. 7, 8-dihydro-8-oxoguanine and 5-hydroxycytosine. *J. Biol. Chem.* **274**, 8516–8523 (1999).
25. Inamdar, K. V. *et al.* Conversion of phosphoglycolate to phosphate termini on 3' overhangs of DNA double strand breaks by the human tyrosyl-DNA phosphodiesterase hTdp1. *J. Biol. Chem.* **277**, 27162–27168 (2002).
26. Bradley, M. O. & Kohn, K. W. X-ray induced DNA double strand break production and repair in mammalian cells as measured by neutral filter elution. *Nucleic Acids Res.* **7**, 793–804 (1979).

27. Kathe, S. D., Shen, G. P. & Wallace, S. S. Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts. *J. Biol. Chem.* **279**, 18511–18520 (2004).
28. Zhou, W. & Doetsch, P. W. Transcription bypass or blockage at single-strand breaks on the DNA template strand: effect of different 3' and 5' flanking groups on the T7 RNA polymerase elongation complex. *Biochemistry* **33**, 14926–14934 (1994).
29. Perry, P. & Wolff, S. New Giemsa method for the differential staining of sister chromatids. *Nature* **251**, 156–158 (1974).
30. Lambert, S. & Lopez, B. S. Role of RAD51 in sister-chromatid exchanges in mammalian cells. *Oncogene* **20**, 6627–6631 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Withers and L. Ju for technical assistance, H. Nash for tyrosine oligonucleotide and critical reading of the manuscript, F. Karimi-Busheri and A. Rasouli-Nia for C-ter3 cells. This work was supported by an MRC Programme Grant to K.W.C. and by an Overseas Research Scholarship and a Stapley Trust award to S.F.E-K.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.W.C. (e-mail: k.w.caldecott@sussex.ac.uk).

Nutrient control of glucose homeostasis through a complex of PGC-1 α and SIRT1

Joseph T. Rodgers¹, Carlos Lerin¹, Wilhelm Haas³, Steven P. Gygi³, Bruce M. Spiegelman^{2,3} & Pere Puigserver¹

¹Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

²Dana-Farber Cancer Institute and ³Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA

Homeostatic mechanisms in mammals respond to hormones and nutrients to maintain blood glucose levels within a narrow range. Caloric restriction causes many changes in glucose metabolism and extends lifespan; however, how this metabolism is connected to the ageing process is largely unknown. We show here that the Sir2 homologue, SIRT1—which modulates ageing in several species^{1–3}—controls the gluconeogenic/glycolytic pathways in liver in response to fasting signals through the transcriptional coactivator PGC-1 α . A nutrient signalling response that is mediated by pyruvate induces SIRT1 protein in liver during fasting. We find that once SIRT1 is induced, it interacts with and deacetylates PGC-1 α at specific lysine residues in an NAD⁺-dependent manner. SIRT1 induces gluconeogenic genes and hepatic glucose output through PGC-1 α , but does not regulate the effects of PGC-1 α on mitochondrial genes. In addition, SIRT1 modulates the effects of PGC-1 α repression of glycolytic genes in response to fasting and pyruvate. Thus, we have identified a molecular mechanism whereby SIRT1 functions in glucose homeostasis as a modulator of PGC-1 α . These findings have strong implications for the basic pathways of energy homeostasis, diabetes and lifespan.

A key regulator of lifespan is the NAD⁺-dependent histone deacetylase Sir2 (silencing information regulator 2), which induces longevity in *Saccharomyces cerevisiae* and *Caenorhabditis elegans* in response to caloric restriction signals^{1–3}. The mammalian homologue, SIRT1—a member of the Sir2 family called sirtuins—is known to target MyoD, p53 and forkhead transcription factors for deacetylation^{4–8}. However, whether and how SIRT1 is involved

in pathways affected directly by caloric restriction in mammals is still unclear.

It is well established that reduced intake of dietary energy results in metabolic changes similar to fasting^{9,10}. Key among these changes are increased fatty-acid oxidation and hepatic gluconeogenesis. PGC-1 α is a key regulator of glucose production in the liver of fasted and diabetic mice through activation of the entire gluconeogenic pathway^{11–15}. In addition to the key hormones insulin, glucagon and glucocorticoids, the rate of hepatic gluconeogenesis is also controlled by nutrients, but how the nutrient response is controlled and whether PGC-1 α is also involved remains unknown.

The fact that Sir2 controls lifespan in *S. cerevisiae* and *C. elegans* in response to caloric restriction^{1,2} prompted us first to investigate whether SIRT1 could also be regulated by nutritional status in mice. As shown in Fig. 1a, *SIRT1* expression in liver was not regulated at the messenger RNA level in fasted mice; however, SIRT1 protein levels were induced after fasting and returned to nearly control levels upon refeeding (Fig. 1b). As previously described¹¹, expression of PGC-1 α was upregulated at the level of mRNA and protein by fasting (Fig. 1a, b). PGC-1 α and SIRT1 increased in correlation with induction of the gluconeogenic gene *PEPCK* (*phosphoenolpyruvate kinase*) (Fig. 1a). We next investigated whether lactate and pyruvate—which can change the ratio of NAD⁺/NADH and regulate SIRT1 activity—were fluctuating in the liver. As shown in Fig. 1c, pyruvate levels were increased 1.7-fold in the liver of fasted mice and conversely lactate levels were decreased twofold. We also measured NAD⁺, a potent activator and substrate of SIRT1. As shown in Fig. 1d, liver NAD⁺ levels were increased by 33% by fasting and returned to control levels after refeeding. Interestingly, we could not detect significant changes of NADH levels (data not shown).

These data suggest that increases of SIRT1 protein levels and NAD⁺ concentration might contribute to the SIRT1 deacetylase activity in the fasted state.

To investigate further the regulation of SIRT1, we examined the effects of cyclic AMP or insulin, key signals of the fasted response in mammals. These two pathways did not affect SIRT1 protein levels in cultured hepatocytes (Supplementary Fig. S1). In contrast, glucose and pyruvate, known to fluctuate in fasting¹⁶, regulated SIRT1 protein levels. As shown in Fig. 1e, increasing concentrations of glucose decreased the amount of SIRT1, whereas increasing concentrations of pyruvate markedly increased it. These changes in SIRT1 protein levels are regulated at the post-transcriptional level as *SIRT1* mRNA remained constant (Fig. 1e). We performed pulse-chase experiments to analyse further whether these changes were at the level of protein synthesis or degradation. As shown in Supplementary Fig. S2, pulsed radiolabelled SIRT1 protein levels did not change after chase for 6 and 12 h under pyruvate treatment. However, an increase in SIRT1 protein synthesis was notably elevated by pyruvate. These results suggest that the increase of SIRT1 protein observed in the liver of fasted mice could be mediated by changes in glucose and/or pyruvate.

The fact that the induction pattern of SIRT1 protein correlated with expression of PGC-1 α in fasting (Fig. 1a, b) suggests that these two proteins could interact. To study this, we performed immunoprecipitation of endogenous SIRT1 protein from liver extracts and precipitated PGC-1 α (Fig. 2a). Moreover, an interaction between both endogenous and/or overexpressed proteins was also observed in cultured hepatocytes (Supplementary Fig. S5b) and 293 cells (Supplementary Fig. S3a). To confirm that this reflected a direct physical interaction, we used an *in vitro* interaction assay with

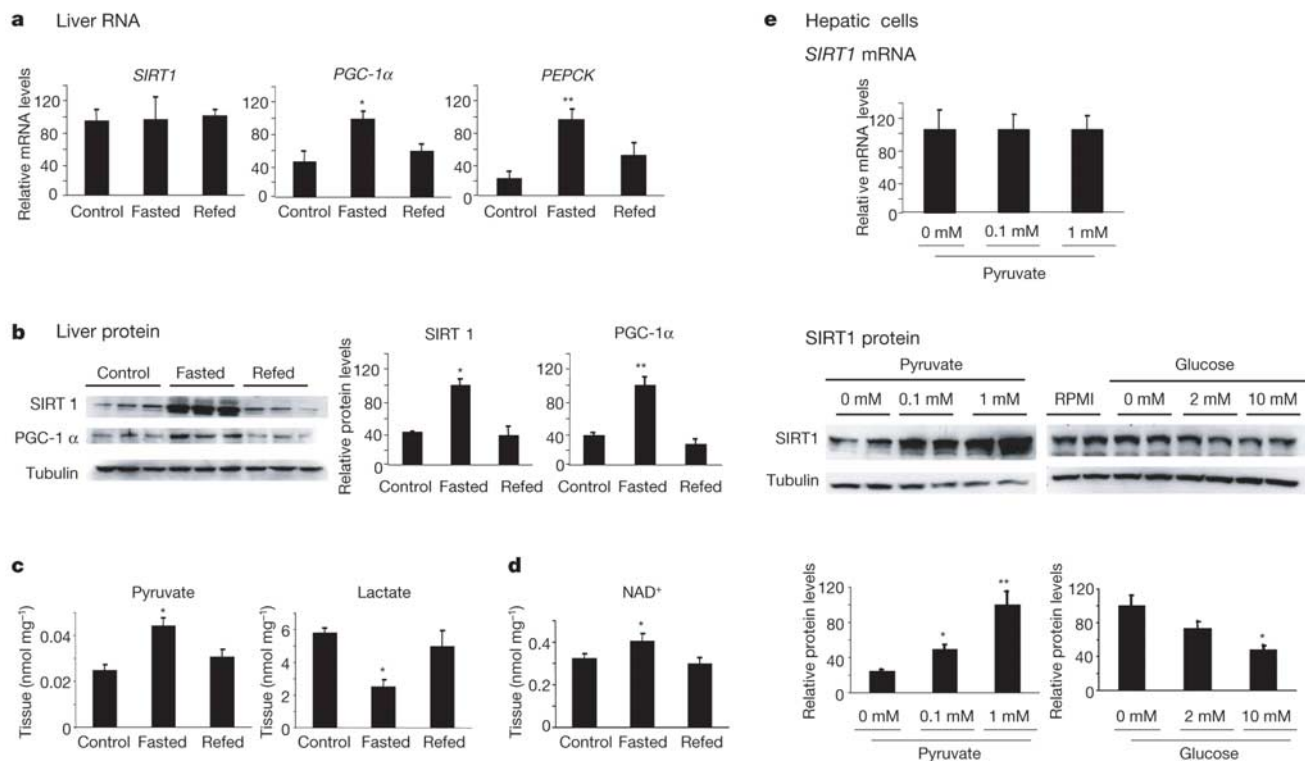


Figure 1 Nutritional regulation of SIRT1 protein and NAD⁺ levels in liver and hepatocytes. **a**, mRNAs for PGC-1 α are induced during fasting in mouse liver. RNA was extracted and Q-RT-PCR analysis was performed in triplicate using specific oligonucleotides (see Supplementary Fig. S8). **b**, SIRT1 protein is regulated by nutritional status. Liver SIRT1 and tubulin protein quantification was performed in duplicate from three samples using Versadoc (Bio-Rad). **c**, Pyruvate and lactate are regulated in the fasted liver. Deproteinized

liver extracts were used to determine lactate and pyruvate. **d**, NAD⁺ levels are increased in the fasted liver. **e**, Glucose and pyruvate control SIRT1 protein levels. Values in **c**, **d** and **e** represent the mean of 3–4 different experiments performed in triplicate. Error bars represent s.e.m. Statistical analyses were performed using Student's *t*-test. **P* < 0.05, ***P* < 0.005.

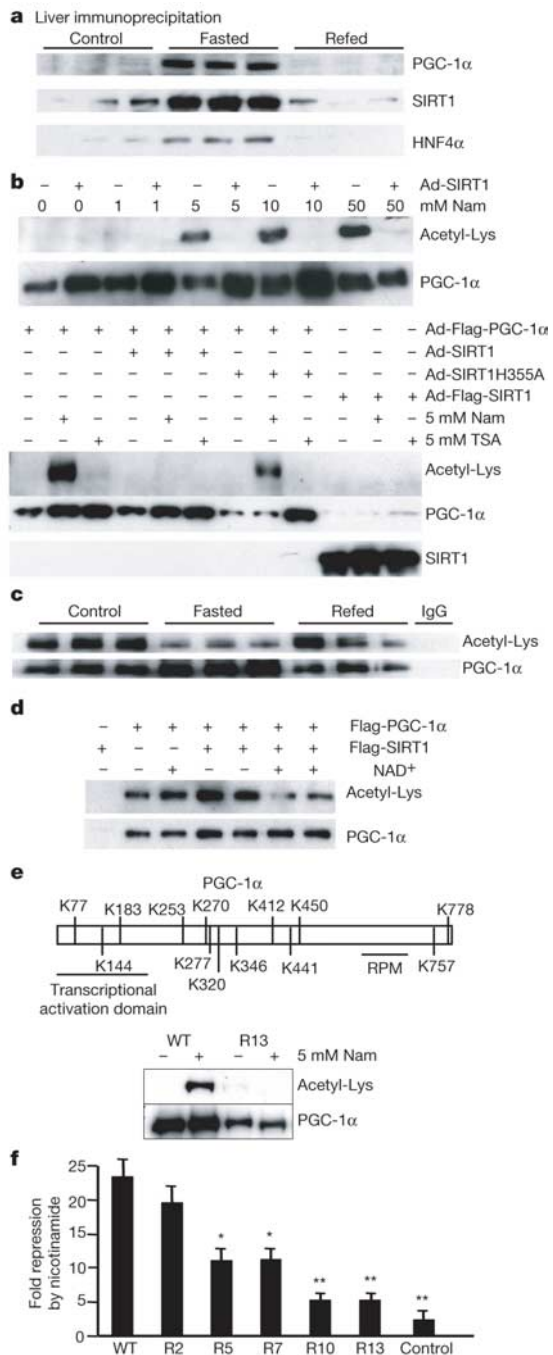


Figure 2 SIRT1 interacts with and deacetylates PGC-1α. **a**, Endogenous SIRT1 interacts with PGC-1α in liver. SIRT1 immunoprecipitation was performed from mouse liver homogenates. Relative input for each lane is shown in Fig. 1b. **b**, SIRT1 deacetylates PGC-1α in cells. 293T cells were treated as in Supplementary Fig. S3a. PGC-1α or SIRT1 were immunoprecipitated and western blot analysis was performed using the indicated antibodies. Nam, nicotinamide; TSA, trichostatin; Ad, adenovirus. **c**, Acetylation of PGC-1α decreases in the fasted liver. Nuclear liver extracts were obtained as described¹⁸. Endogenous PGC-1α was immunoprecipitated as described in the Methods section. IgG antibodies were used as a negative control. **d**, NAD⁺-dependent deacetylation of PGC-1α by SIRT1. **e**, PGC-1α lysine acetylation sites. 293T cells were infected with adenoviruses expressing Flag-PGC-1α or Flag-PGC-1α R13. Acetylated lysine residues were identified by tandem mass spectrometry (Supplementary Fig. S4). **f**, Nicotinamide represses the transcriptional activity of PGC-1α on HNF4α. 293 cells were transfected with plasmids encoding an HNF4α binding site luciferase, HNF4α and PGC-1α wild type, R2 (K183R, K253R), R5 (R2 + K320R, K346R, K441R), R7 (R5 + K450R, K778R), R10 (R7 + K77R, K412R, K757R) and R13 (R10 + K144R, K270R, K277R). Values represent means of four different experiments performed in duplicate; error bars represent s.e.m. **P* < 0.05; ***P* < 0.005, control versus nicotinamide.

recombinant glutathione S-transferase (GST)–PGC-1α (Supplementary Fig. S3b). The fact that PGC-1α and SIRT1 directly interacted suggests that PGC-1α might be a substrate for SIRT1 deacetylation. As shown in Fig. 2b, nicotinamide (SIRT1 inhibitor) treatment strongly induced PGC-1α lysine acetylation. Importantly, expression of SIRT1 could overcome nicotinamide-induced PGC-1α acetylation, but a SIRT1 mutant (SIRT1H355A) that lacks enzymatic activity had no effect (Fig. 2b). PGC-1α acetylation decreased in the fasted liver, suggesting that SIRT1 activity is increased in this situation (Fig. 2c). To demonstrate that SIRT1 deacetylated PGC-1α directly, we used an *in vitro* deacetylase assay. As shown in Fig. 2d, PGC-1α acetylation is decreased (60%) only upon addition of both SIRT1 and NAD⁺. To map the PGC-1α lysine-acetylation sites induced by nicotinamide, tandem mass spectrometry analysis was performed. We found that PGC-1α was acetylated at 13 lysine sites (Fig. 2e and Supplementary Fig. S4). Mutation of these 13 lysines to arginine abolished the acetylation of PGC-1α (Fig. 2e). To test the effect of PGC-1α acetylation on its transcriptional activity, luciferase reporter assays using the transcription factor HNF4α were performed. As shown in Fig. 2f, nicotinamide treatment decreased the transcriptional activity of wild-type PGC-1α/HNF4α by 24-fold. We tested several combinations of PGC-1α mutants and found that a mutant of PGC-1α (R5), in which five lysines were substituted to arginine, decreased the fold repression by nicotinamide to tenfold. In addition, two different mutants R10 and R13, containing the R5 mutations, largely decreased this repression to almost control levels (Fig. 2f). These results suggest that acetylation of PGC-1α decreases its ability to efficiently coactivate HNF4α. The molecular mechanisms underlying this transcriptional repression by acetylation are under current investigation.

Previous studies have focused on the hormonal regulation of PGC-1α expression^{11–13}. However, the above data prompted us to investigate whether the gluconeogenic function of PGC-1α might also be controlled through a nutrient-sensing pathway via pyruvate and SIRT1. As shown in Fig. 3a, the gluconeogenic genes *PEPCK* and *G6Pase* (*glucose-6-phosphatase*) were induced by pyruvate treatment. The effect of pyruvate on the induction of *G6Pase* and *PEPCK* was decreased by a specific short interfering RNA for PGC-1α (*G6Pase* decreased by 30% and *PEPCK* by 70%; Fig. 3a) but not by a scrambled siRNA (control); this indicates that the pyruvate regulation of these genes required PGC-1α. We next used siRNA for *SIRT1* (Supplementary Fig. S6) to determine whether the effect of pyruvate on PGC-1α was dependent on SIRT1. Indeed, this knock-down of SIRT1 protein largely blocked the pyruvate-induced PGC-1α increase on *PEPCK* and *G6Pase* (Fig. 3b). The pyruvate and SIRT1 effects were specific to gluconeogenic genes because PGC-1α-targeted mitochondrial genes (*cytochrome-c* and *β-ATP synthase*) were not significantly changed. These data indicate a requirement of SIRT1 in the regulation of gluconeogenic gene expression by PGC-1α. Consistent with these effects, endogenous SIRT1 and PGC-1α were recruited to chromatin promoter fragments of the PGC-1α targets *PEPCK* and *G6Pase*. Notably, the recruitment of both proteins was increased by pyruvate treatment (Supplementary Fig. S7). PGC-1α requires interaction with HNF4α to induce gluconeogenic gene expression¹⁴. We therefore examined whether SIRT1 was in a complex with HNF4α and PGC-1α. As shown in Fig. 2a and Supplementary Fig. S5, endogenous or over-expressed HNF4α coprecipitated with SIRT1 and PGC-1α. Taken together, these results might suggest that PGC-1α and SIRT1 are in a protein complex that includes HNF4α.

The metabolic flux of glucose in hepatocytes is also affected by the rate of glycolysis, a pathway that is decreased during fasting. To test whether PGC-1α and SIRT1 might also control glycolysis, we analysed the expression of *glucokinase* and *liver pyruvate kinase* (*LPK*). As shown in Fig. 3a, b, PGC-1α and pyruvate decreased the levels of mRNAs for *glucokinase* and *LPK*. *SIRT1* siRNA increased

the expression of these genes under pyruvate treatment. These results suggest that PGC-1 α and SIRT1 co-regulate, in opposite directions, the patterns of gluconeogenic and glycolytic gene expression in response to pyruvate. We next investigated whether this induction of gluconeogenic and glycolytic genes by SIRT1 and PGC-1 α was reflected in glucose production *per se*. As shown in Fig. 4a, PGC-1 α increased glucose production threefold compared with control cells. *SIRT1* siRNA repressed the glucose production induced by PGC-1 α to control levels. This indicates that SIRT1 is required for PGC-1 α induction of glucose production.

The data we present here show for the first time that PGC-1 α and SIRT1, which have been studied in very different contexts, can function together to promote adaptation to nutrient deprivation by regulating the genetic programmes of gluconeogenesis and glycolysis. This provides one of the most striking examples of a biological programme regulated through transcriptional coactivation (Fig. 4b).

SIRT1 is in a protein complex with PGC-1 α and HNF4 α , an

essential transcription factor in PGC-1 α 's gluconeogenic function¹⁴. In this protein complex SIRT1 is likely to be the sensor for nutrient fluctuations via NAD⁺ and regulates PGC-1 α -dependent gene expression. Whether other PGC-1 α gluconeogenic interacting transcription factors that are hormonally regulated, such as FOXO1 and glucocorticoid receptor, are also involved in SIRT1-mediated effects is unknown. SIRT1 has mainly been linked to negative regulation of gene expression through protein deacetylation^{5–7}. However, we show here that SIRT1 can act both positively and negatively to control gene expression as a cofactor for PGC-1 α . Although the mechanism of this transcriptional regulation is unknown, it is possible that the recruitment of a different set of coactivators and corepressors through PGC-1 α /SIRT1 could explain these opposite effects. Interestingly, it has recently been reported that SIRT1 interacts with the transcriptional corepressor NCoR, negatively regulating PPAR γ in white fat where PGC-1 α is very low¹⁷.

SIRT1 has been implicated in other specific biological responses, such as the reaction to oxidative stress through deacetylation of p53

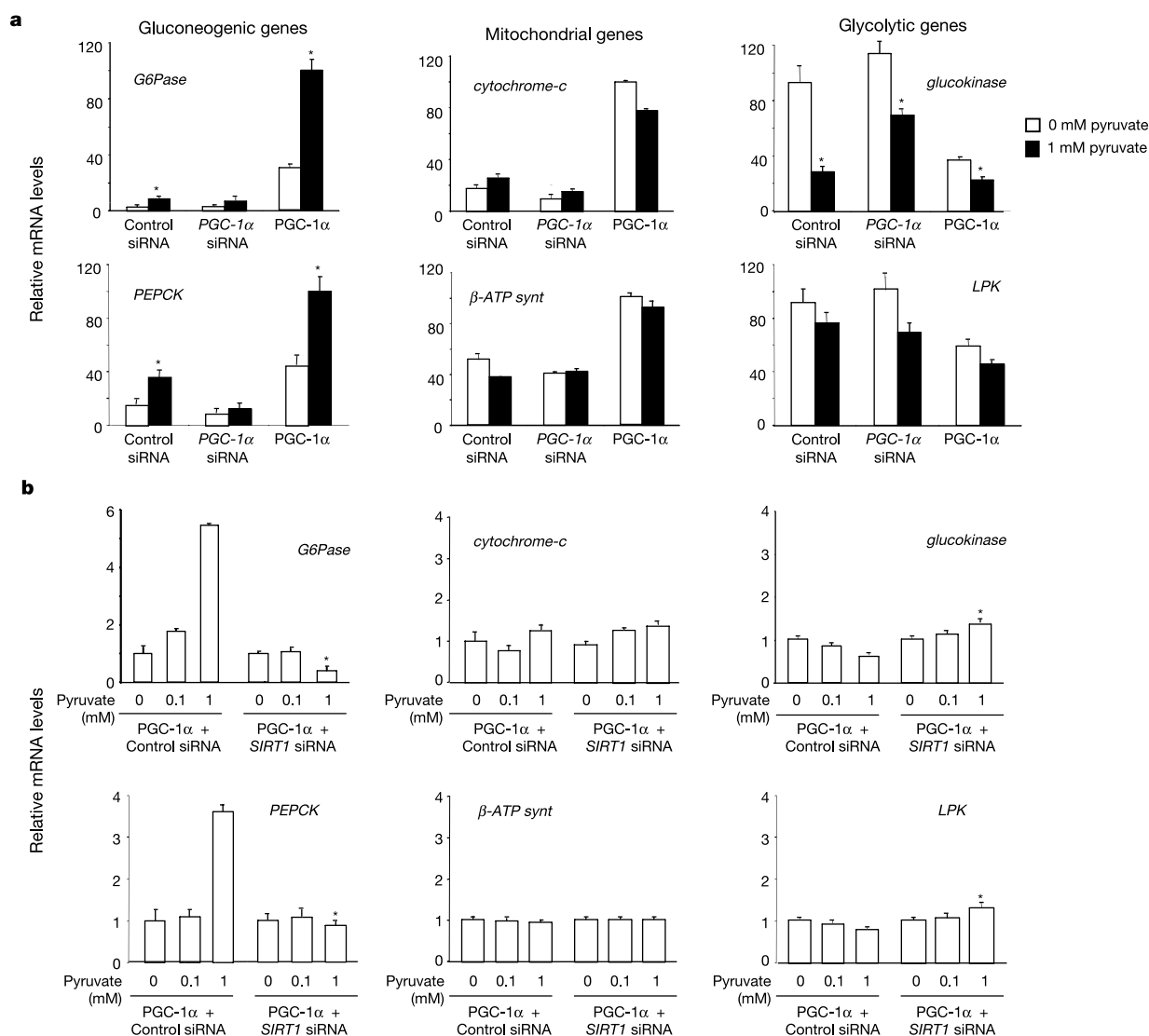


Figure 3 Nutrient regulation of gluconeogenic and glycolytic genes is controlled through SIRT1. **a**, Pyruvate and PGC-1 α increase gluconeogenic and decrease glycolytic gene expression. Fao hepatocytes were infected with the indicated adenoviruses and incubated with 1 mM pyruvate for the last 6 h. The mRNA values represent the average of three different experiments made in duplicate and error bars represent the s.e.m. * $P < 0.05$,

untreated versus pyruvate treatment. **b**, SIRT1 is required for the pyruvate and PGC-1 α regulation of gluconeogenic and glycolytic genes. The mRNA values represent the average of four different experiments made in duplicate and error bars represent the s.e.m. * $P < 0.01$, control siRNA versus *SIRT1* siRNA under 1 mM pyruvate treatment.

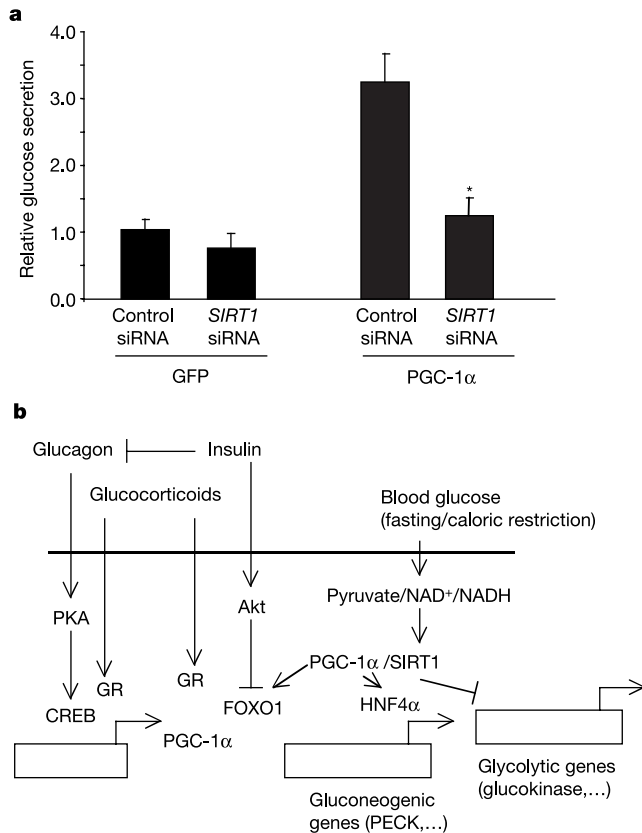


Figure 4 Nutrient regulation of PGC-1 α function on glucose output depends on SIRT1. **a**, Fao hepatocytes were infected with the indicated adenoviruses. Values represent the mean of three different experiments made in duplicate and error bars represent the s.e.m. * $P < 0.01$, control versus SIRT1 siRNA. Levels for PGC-1 α and SIRT1 are shown in Supplementary Fig. S6. **b**, Model illustrating the hormonal and nutrient regulation of gluconeogenic and glycolytic gene expression. GR, glucocorticoid receptor; CREB, cyclic AMP response element binding protein; PKA, protein kinase A; PECK, phosphoenol pyruvate kinase; Akt, PKB/Akt).

and FOXO1 transcription factors^{5–8}. Whether PGC-1 α is also involved in this oxidative stress response is an interesting question because PGC-1 α interacts with FOXO1 in hepatocytes¹³. It is therefore possible that multiple effects of SIRT1 in energy metabolism or even lifespan might occur by means of docking on this coactivator. Ultimately, a better understanding of these nutritional signalling systems may lead to new therapeutic approaches to defects in glucose metabolism in diabetes and ageing. □

Methods

Constructs

Wild-type and H355A SIRT1, and control and SIRT1 siRNA adenoviruses were constructed using the pAd-Easy system. Inserts were cloned into the pAdTrack-CMV shuttle vector, and adenovirus constructs were created by recombination of the shuttle vector and pAdEasy vector by electroporation into BJ5183-AD-1 bacteria (Stratagene).

Cell culture and treatments

Fao rat hepatocytes were cultured in RPMI with 10% fetal bovine serum (FBS). Treatments with nicotinamide (5 mM), insulin (10 nM) and forskolin (1 μ M) (Sigma) were performed in 0.5% bovine serum albumin (BSA) for 6 h. Glucose and pyruvate treatments were performed for 6 h in distilled PBS (Gibco) with 0.5% BSA.

Animal experiments

Four-week-old C57B1/6 mice were fed *ad libitum*, fasted for 24 h and refed for 24 h. Animals were killed and the livers were removed and snap-frozen. Liver whole-cell homogenates were made with RIPA buffer and used for western blot analysis and metabolite measurements. Liver mRNA was isolated using Trizol reagent (Invitrogen) and used for northern blot analysis and real-time polymerase chain reaction (PCR).

Protein interaction analysis

Endogenous protein interactions from liver tissues were performed by co-immunoprecipitation experiments using a SIRT1 and PGC-1 α antibody. Liver extracts were obtained by homogenization with a buffer containing 20 mM HEPES-KOH (pH 7.9), 125 mM NaCl, 0.1% NP40, 1 mM EDTA, 0.1% Triton-X-100, 5 mM nicotinamide, 5 μ M TSA, protease and phosphatase inhibitors. Protein–protein interactions in cells were also performed by co-immunoprecipitation experiments. Flag-tagged proteins were expressed in 293 cells using Superfect (Qiagen) or adenoviral infection. Twenty-four hours after transfection and/or infection, whole-cell extracts were prepared and subjected to immunoprecipitation with Flag antibody (Sigma) linked to agarose beads. The immunoprecipitates were separated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) and immunoblotted using antibodies against SIRT1 (Upstate Biotechnology), HNF4 α (Santa-Cruz), PGC-1 α (Santa-Cruz) and tubulin (Upstate Biotechnology).

PGC-1 α acetylation assays

PGC-1 α lysine acetylation was analysed by immunoprecipitation of PGC-1 α followed by western blot using acetyl-lysine antibodies (Cell Signalling and Technology). Liver nuclear extracts were obtained as described¹⁸. PGC-1 α was immunoprecipitated using H-300 antibody from Santa-Cruz. PGC-1 α levels and acetylation were detected using specific antibodies for PGC-1 α and acetyl lysine. Details on mapping PGC-1 α acetylation sites are provided in Supplementary Information S4.

In vitro deacetylation assay

293 cells were infected with either Flag-PGC-1 α adenovirus and treated with 5 mM nicotinamide, or Flag-SIRT1 adenovirus separately. After 24 h the cells were collected in RIPA buffer (1% NP-40, 0.1% SDS, 0.5% NaDeoxycholate in PBS), immunoprecipitated and eluted using Flag peptide. The elutant Flag-PGC-1 α and Flag-SIRT1 were combined as indicated in reaction buffer (50 mM Tris-HCl pH 9.0, 50 mM NaCl, 4 mM MgCl₂, 0.5 mM DTT) as described⁵, with or without the addition of 100 μ M NAD⁺, for 4.5 h at 30 °C.

Transcriptional activation assays

293 cells were transiently transfected in six-well dishes using Superfect (Qiagen). The ratio DNA:Superfect was 1:1.5. Cell culture medium was changed after 12 h of transfection. After 24 h, cells were treated with nicotinamide for 12 h and luciferase assays were performed.

Gene expression analysis

Total RNA prepared from either Fao cells or from liver was extracted with Trizol (Invitrogen). Complementary DNA generated by Superscript II enzyme (Invitrogen) was analysed by quantitative reverse-transcriptase-mediated PCR (Q-RT–PCR) using an iQ SYBR Green Supermix (Bio-Rad). All data were normalized to tubulin expression. The oligonucleotide primers used are provided as Supplementary Information (Supplementary Fig. S8). In some experiments, northern blot analysis was also performed. Expression of mRNA was quantified using a Bio-Rad Personal Molecular Imager FX, and quantification was performed using Quantity One quantification software (Bio-Rad).

Fao adenovirus infection

Fao cells were infected with adenovirus for 24 h in RPMI with 0.5% BSA. The media was then replaced with fresh RPMI with 0.5% BSA for an additional 24 h. The cells were then washed with PBS and incubated with dPBS (Gibco) with 0.5% BSA for 6 h containing treatment as indicated. Cells were either collected with RIPA buffer for protein analysis or Trizol (Invitrogen) for RNA analysis.

Hepatic glucose output

Fao cells were infected with adenoviruses expressing PGC-1 α and SIRT1 siRNA. Twenty-four hours after infection the medium was replaced with 1 ml of glucose-free DMEM, supplemented with 20 mM sodium lactate and 2 mM sodium pyruvate. After a 6-h incubation, 0.5 ml of medium was collected and the glucose concentration measured using the glucose phosphate dehydrogenase method¹⁹. The readings were then normalized to the total protein content.

NAD⁺ and NADH measurements

Fao cells and liver tissue were deproteinized and the concentrations of lactate and pyruvate were determined as previously described^{19,20}. NAD⁺ and NADH nucleotides were directly measured as described^{21,22}. In brief, approximately 20 mg of frozen liver tissues were homogenized in 200 μ l of acid extraction buffer to obtain NAD⁺ concentration, or 200 μ l of alkali extraction to obtain NADH concentration. Homogenates were neutralized with 100 μ l of buffer. The concentration of nucleotides was measured fluorimetrically after an enzymatic cycling reaction using 5 μ l of sample. Values for both nucleotides were detected within the linear range.

Received 4 November 2004; accepted 6 January 2005; doi:10.1038/nature03354.

- Kaeblerlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).
- Tissenbaum, H. A. & Guarente, L. Increased dosage of a sir-2 gene extends lifespan in *Caenorhabditis elegans*. *Nature* **410**, 227–230 (2001).
- Smith, J. S. *et al.* A phylogenetically conserved NAD⁺-dependent protein deacetylase activity in the Sir2 protein family. *Proc. Natl Acad. Sci. USA* **97**, 6658–6663 (2000).
- Fulco, M. *et al.* Sir2 regulates skeletal muscle differentiation as a potential sensor of the redox state. *Mol. Cell* **12**, 51–62 (2003).

5. Luo, J. *et al.* Negative control of p53 by Sir2 α promotes cell survival under stress. *Cell* **107**, 137–148 (2001).
6. Vaziri, H. *et al.* hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase. *Cell* **107**, 149–159 (2001).
7. Motta, M. C. *et al.* Mammalian SIRT1 represses forkhead transcription factors. *Cell* **116**, 551–563 (2004).
8. Brunet, A. *et al.* Stress-dependent regulation of FOXO transcription factors by the SIRT1 deacetylase. *Science* **303**, 2011–2015 (2004).
9. Dhahbi, J. M. *et al.* Calories and aging alter gene expression for gluconeogenic, glycolytic, and nitrogen-metabolizing enzymes. *Am. J. Physiol.* **277**, E352–E360 (1999).
10. Hagopian, K., Ramsey, J. J. & Weindrich, R. Caloric restriction increases gluconeogenic and transaminase enzyme activities in mouse liver. *Exp. Gerontol.* **38**, 267–278 (2003).
11. Yoon, J. C. *et al.* Control of hepatic gluconeogenesis through the transcriptional coactivator PGC-1. *Nature* **413**, 131–138 (2001).
12. Herzig, S. *et al.* CREB regulates hepatic gluconeogenesis through the coactivator PGC-1. *Nature* **413**, 179–183 (2001).
13. Puigserver, P. *et al.* Insulin-regulated hepatic gluconeogenesis through FOXO1-PGC-1 α interaction. *Nature* **423**, 550–555 (2003).
14. Rhee, J. *et al.* Regulation of hepatic fasting response by PPAR γ coactivator-1 α (PGC-1): requirement for hepatocyte nuclear factor 4 α in gluconeogenesis. *Proc. Natl Acad. Sci. USA* **100**, 4012–4017 (2003).
15. Lin, J. *et al.* Defects in adaptive energy metabolism with CNS-linked hyperactivity in PGC-1 α null mice. *Cell* **119**, 121–135 (2004).
16. MacDonald, M., Neufeldt, N., Park, B. N., Berger, M. & Ruderman, N. Alanine metabolism and gluconeogenesis in the rat. *Am. J. Physiol.* **231**, 619–626 (1976).
17. Picard, F. *et al.* Sirt1 promotes fat mobilization in white adipocytes by repressing PPAR- γ . *Nature* **429**, 771–776 (2004).
18. Timchenko, N. A., Wilde, M. & Darlington, G. J. C/EBP α regulates formation of S-phase-specific E2F-p107 complexes in livers of newborn mice. *Mol. Cell. Biol.* **19**, 2936–2945 (1999).
19. Williamson, D. H., Lund, P. & Krebs, H. A. The redox state of free nicotinamide-adenine dinucleotide in the cytoplasm and mitochondria of rat liver. *Biochem. J.* **103**, 514–527 (1967).
20. Zhang, Q., Yao, H., Vo, N. & Goodman, R. H. Acetylation of adenovirus E1A regulates binding of the transcriptional corepressor CtBP. *Proc. Natl Acad. Sci. USA* **97**, 14323–14328 (2000).
21. Lin, S. S., Manchester, J. K. & Gordon, J. I. Enhanced gluconeogenesis and increased energy storage as hallmarks of aging in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**, 36000–36007 (2001).
22. Lin, S. J., Ford, E., Haigis, M., Liszt, G. & Guarente, L. Calorie restriction extends yeast life span by lowering the level of NADH. *Genes Dev.* **18**, 12–16 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Vazquez for important discussions on the project. We also acknowledge members of the Puigserver laboratory for helpful comments on this work, especially T. Cunningham and S.-H. Kim for technical assistance. Some constructs or reagents were obtained from W. Gu, L. Guarente, D. Robinson and M. Stoffel. We also thank M. Montminy for the PGC-1 α polyclonal antibody. The protocols for NAD⁺ and NADH measurements were obtained from S.-J. Lin. Part of these studies was supported by awards from the Ellison Medical Foundation and the American Federation for Aging Research (P.P.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.P. (ppuigse1@bs.jhmi.edu).

corrigendum

Drosophila dFOXO controls lifespan and regulates insulin signalling in brain and fat body

Dae Sung Hwangbo, Boris Gershman, Meng-Ping Tu, Michael Palmer & Marc Tatar

Nature **429**, 562–566 (2004).

In this Letter, Boris Gershman's surname was misspelled as 'Gersham'. It is presented correctly here. □

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayouni (4978)

European Sales Manager:

Andy Douglas (4975)

UK/ RoW/ Ireland:

Nils Moeller (4953)
Irene Vigila-Alton (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents:

Sille Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Advertising Production

Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Niamh Shields

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan

Tel + 49 89 54 90 57 11

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakata-nachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

Fighting urban myths

Pets flushed down toilets that turn into monster-sized, sewer-prowling predators. Large companies that promote occultism through secret messages in the labels on their products. Popular consumer goods that actually have toxic properties. All these are examples of urban myths — tales that have circulated by either word of mouth or forwarded e-mails. Many of these stories, and similar examples, perch perpetually on the threshold of belief, perhaps because they exploit people's fears so well. As a result, they tend to persist despite being frequently debunked.

Scientific employment has its own urban myths. PhDs who end up pursuing a career as a taxi driver. Legions of postdocs mired in fellowships that last a lifetime. The need for ever-more scientists in a particular 'hot' subdiscipline, despite an abundant supply of scientists in general.

Such myths can only be fought effectively if you have an abundance of reliable data, a panel of experts told a careers workshop at the annual meeting of the American Association for the Advancement of Science in Washington DC last week. But good information is not always easy to find — at least not all in one place. And even the best data on scientific employment and the workforce tend to be too old and broad to produce good projections for specific parts of the world or for narrow subdisciplines.

But this is set to change, the panellists said. The Commission on Professionals in Science and Technology (www.cpst.org) is bringing together key data from the US Bureau of Labor Statistics, US census data, the National Science Foundation and a host of other scientific organizations. The new database should make it easier for people to debunk any scientific-career myths standing in the way of their professional path. For scientists wishing to confront such issues in their own workplace, the truth is out there.

Paul Smaglik

Naturejobs editor



Contents

CAREER VIEW

Nuts & Bolts

Networking for introverts

Graduate Journal

A hard day's night

Movers

Deepak Srivastava

p120

WWW.NATUREJOBS.COM

Career centre

Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

GRADUATE JOURNAL

A hard day's night

What is it that makes me get up early in the morning after too little sleep? When the previous day was spent wrestling with another failed experiment, and cutting short phone calls home for another bunch of articles that needed to be read? It's strange, but I can't answer straight away.

Making science work — even at a basic level like in my graduate research projects — is attractive to much more than my intellect. The lab is the place where I spend most of the day and part of the night as well. My colleagues are the first people I share my thoughts with. It's hard to explain this level of engagement and involvement to others outside my little research universe. But sometimes it's pivotal to face the truth: doing research converts me into an egoist who chases ideas and results, and thereby forgets about time and society.

Despite the intensity of my lab life, it is a relief to have some people outside who I can't talk to about the stuff that occupies me the whole day. They wouldn't understand a word about what I spend most of my time doing. So in the end it's this ignorance that saves me and makes my day. It's this small preserve outside science that gives me time to recharge and get up in the morning to start off to the lab for another long day. ■

Tobias Langenhan is a first-year graduate student in neuroscience at the University of Oxford, UK.

NUTS & BOLTS

Networking for introverts

For most scientists, spending solitary hours pursuing research is a common and comfortable endeavour. So, what to do about the suggestion that networking is a 'must do' activity for a successful career progression? Surely there must be a way around it?

The way around it is through it. Like their business counterparts, professionals in the research community need to shape their work lives, create their own opportunities and build relationships to accomplish their goals. In this process, introverts are sometimes undervalued and overlooked. With an estimated 75% of job offers and the majority of career advancement resulting from some form of networking, it's a role too central to success to disregard.

The key to networking is to find an approach you



With Deb Koen
Careers consultant

can follow in your own way and in your own time.

Adapting the process to suit your style, while adopting practices that alleviate the associated stress, will allow you to get the most out of networking. Introversion can actually be a benefit, when applied judiciously. Your conciseness and clarity will be highly prized, particularly by people who have been subjected to an extrovert's seeming disregard for time and boundaries.

In large group networking events, a practised introduction will pave the way for smooth

encounters. Remember to make eye contact and prepare written scripts beforehand to ease your way through conversations. Your audience will appreciate the active listening skills that tend to come naturally to introverts. Following your strengths and trusting the process will allow you to be more in control, to make new discoveries and to boost your confidence.

Celebrate small successes whenever you push yourself out of your comfort zone. The magic of networking is that although you can't predict exactly how or when your efforts will bear fruit, there will be moments of insight and times of connection. Sometimes the people who have no obvious connection to your career end up being especially helpful and taking a sincere interest in your progress. ■

Deb Koen is vice-president of Career Development Services and a columnist for The Wall Street Journal's CareerJournal.com.

MOVERS

Deepak Srivastava, director, Gladstone Institute of Cardiovascular Disease, San Francisco



An overnight sailing trip helped Deepak Srivastava to plot a course for his scientific career.

Brad Thompson, his mentor at medical school, took the newly minted MD out on his boat to celebrate Srivastava's graduation.

The dark calm of the night sea provided Srivastava with an opportunity to chart his future. "It was one of those

discussions you have with no lights or no electricity," Srivastava says. "I asked Brad if he were a young hotshot, which direction would he take."

Thompson replied that looking at pluripotent cells — cells that can differentiate into many other types — would be a major growth area. His prophecy would ring true years later when human embryonic stem cells were cultivated.

Srivastava experienced another moment of clarity during a stint as a paediatric resident at the University of California, San Francisco. "I really enjoyed taking care of infants with heart disease," Srivastava says. "The cells didn't get told to do the right things, so it took me back to my initial research interest."

Those two moments of insight combined to lead Srivastava into a research training programme for paediatric scientists. This allowed him

to focus on the lab, with no need for teaching, grant-writing or making clinical rounds.

Since then, Srivastava has primarily worked on how precursor cells become heart cells — first based in Boston, then at the MD Anderson Cancer Center in Houston, Texas. At the latter, he met Eric Olson, another key mentor with whom he has continued to collaborate after they both moved to the University of Texas Southwestern Medical Center.

At the helm of the Gladstone Institute of Cardiovascular Disease at the University of California, San Francisco, Srivastava will be able to take a genetic and developmental-biology approach to heart disease, including examining the role of gene regulation and cell differentiation in heart disease. San Francisco is an especially good place to tackle these problems, he says — especially as California recently voted to fund stem-cell research. ■

CV

1996–2005: University of Texas Southwestern Medical Center, Dallas, Texas (rising to professor, departments of paediatrics and molecular biology).

1996–2005: Attending physician, paediatric cardiology, Children's Medical Center of Dallas, Texas.

1994–96: Fellow, MD Anderson Cancer Center, Houston, Texas.

1992–94: Fellowship, cardiology department, Children's Hospital, Boston, Massachusetts.

1990–92: Resident, Department of Pediatrics, University of California, San Francisco.

A modest proposal...

...for the perfection of nature.

Vonda N. McIntyre

The crop grows like endless golden silk. Wave after wave rushes across plains, between mountains, through valleys, in a tsunami of light.

Its harvest is perfection. It fills the nutritional needs of every human being. It adapts to our tongues, creating the taste, texture and satisfaction of comfort food or dessert, crisp vegetables or icy lemonade, sea cucumber or big game. It's the pinnacle of the genetic engineer's art.

It's the last and only living member of the plant kingdom on Earth.

Solar cells cover slopes too steep and peaks too high for the monoculture. The solar arrays flow in long, wide swaths of glass, gleaming with a subtle iridescence, collecting sunlight. Our civilization never runs short of power.

The flood of grain drowns marsh and desert, forest and plain, bird and beast and insect. Land must serve to produce the crop; creatures only nibble and trample and damage it, diverting resources from the service of human beings. Even the immortality of rats and cockroaches has failed.

The grain stops at the ocean's beach. No rivers muddy the sea's surface or break the shoreline. The grain and the cities require fresh water, and divert it before it wastes itself in the sea.

The tides wash up and back, smoothing the clean silver sand, leaving it bare of tangled seaweed, of foraging seabirds or burrowing clams, of the brown organic froth that dirtied it in earlier times. Now and then the waves erase a line of human footprints, but these are very rare.

The air is clear of any bite of iodine, any hint of pollution or decay.

The sea undulates, blue and green, clear as new glass. Sunlight shimmers on its surface and dapples the bare sea floor. Underwater turbines cast shadows on the sand. The tides power the turbines, tapping the force of gravity.

Far from shore, where its colonies will not interrupt the vista of clear water, a single species of cyanobacterium photosynthesizes near the surface, pumping oxygen into the crystalline air, controlling the level of carbon dioxide. Its design copes easily with the increasing saltiness of the sea.

Except for the cyanobacteria, the ocean's cacophony of microscopic organisms has followed redwoods, mammoths and *Hallucigenia* into extinction. The krill are gone. Krill would be of as little use to people as sharks and seabirds, fish or jellyfish, seashells or whales. They are all gone, too.



The water deepens beyond the reach of light. The continental shelf ends in a precipice, dropping off into darkness.

On the sea floor, the glass-lace shells of diatoms lie clean and dead, slowly settling. In a moment of geologic time, they will form white limestone.

In the deepest trenches, black smokers gush scalding chemical soup. Machines sense the vents of heat, swim to them, and settle over them to trap the energy from the centre of Earth. Nothing remains for the sustenance and evolution of primordial life in these extraordinary environments.

The strange creatures that lived there, and died, were never any use to human beings.

All the resources of sea and land serve our needs.

Cities of alabaster and adamantine grace the crests of mountains and span the flow of rivers. The cities' people live rich, full lives, long and healthy, free of disease. We are well fed. We have interesting, challenging occupations and plenty of time for leisure, family and virtual reality. We can experience any adventure, from wilderness to exotic ritual, without the expense, trouble or danger of travel. We can experience any adventure that ever happened, any adventure anyone can imagine. The virtual experience matches reality or invention in every way:

sight, sound, smell, touch and movement.

Our civilization pulses with vitality. We have unlimited opportunity: of thought, of achievement, of freedom, and of the pursuit of happiness.

Whatever we require, human ingenuity can invent and provide. And if, in some unlikely but imaginable future, we should wish to recreate any organism, the means to do so exist. DNA sequences, RNA sequences, are easy to write down and archive; there is no need to store messy biological material, either tough and persistent DNA or fragile and degradable RNA. We are magnanimous; we have preserved the blueprints for everything, even parasites and pathogens.

No one has bothered to recreate an organism in a very long time. We have considered the question long and hard, and we have made our decision. No creation of nature has an inherent right to exist, independent of our need.

We have perfected nature, for we are its masters. ■

Vonda N. McIntyre's work runs from Louis XIV (The Moon and the Sun) to space (Little Faces, SciFiction 2005), from local television (producer-director of Science Fiction Conversations) to ancient Crete (The Curve of the World, in progress). Learn more at www.vondanmcintyre.com, www.sfconversations.com and www.scifi.com/scifiction.